ExPlain Transition Overview

The current ExPlain tool will be phased out at the end of 2015, replaced with an enhanced version of TRANSFAC. This document provides a mapping of the functions in ExPlain to the comparable functions in TRANSFAC to help you transition to the new interface.

While many of the functions of ExPlain have been integrated into TRANSFAC and enhanced further, the set of tools that are primarily focused on basic statistical processing and conversion of raw data files have not been migrated due to low usage combined with the ready accessibility of high quality data pre-processing tools made available by instrument providers and public resources.

For a more complete description of the features included in this document, as well as a complete description of other features included in TRANSFAC, please access the TRANSFAC user manual using the help menu accessible in the type right corner of the TRANSFAC interface:

For any questions, or to request that additional new features be considered for future implementation, please contact us at support@biobase-international.com.

Mapped features

Loading gene and miRNA sets

Creating data subsets

Loading sequences and intervals

Match and Composite model search

Loading matrices

Creating profiles

Creating composite models

Last updated September 30, 2015

[Finding novel motifs (Seeder)]

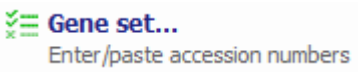
[Functional analysis]

[Network analysis]

[Step-by-step analysis (Wizard workflows)]

Loading gene and miRNA sets

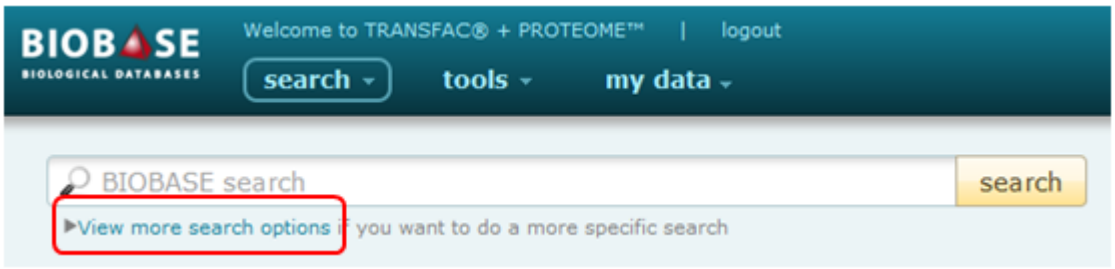
The options described here replace the '**File -> Create new data -> Gene set**' and '**File -> Load data from file -> Gene set**' functions in ExPlain:



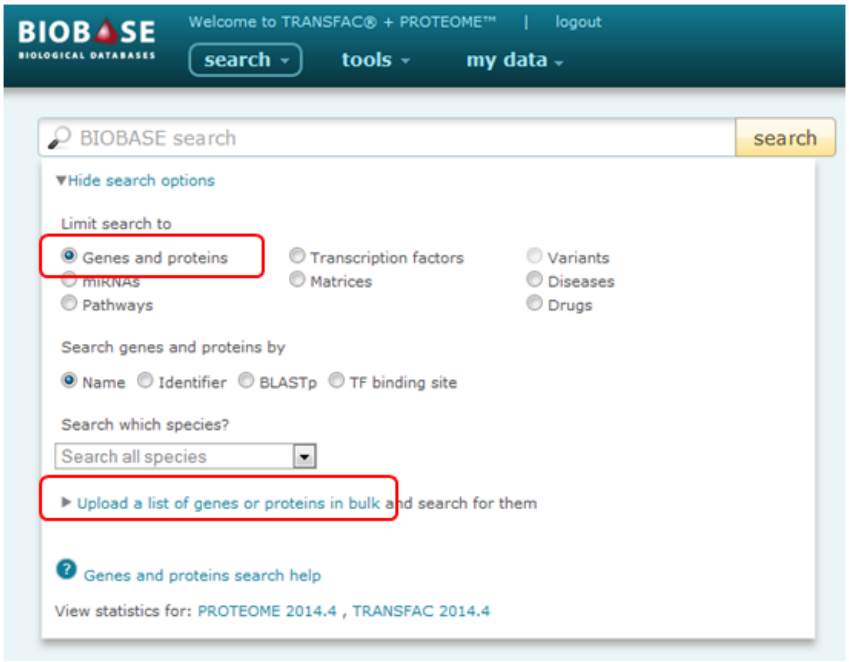
There are two options for loading a list of genes into TRANSFAC for analysis by Match and other tools:

1. Genes and proteins search

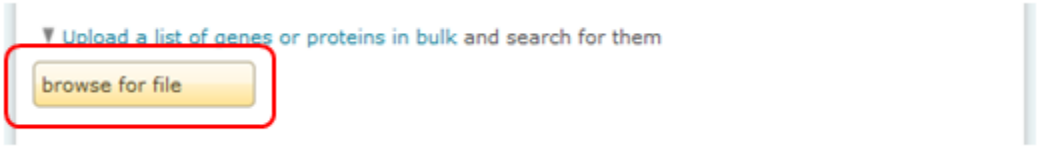
From the TRANSFAC home page (click the BIOBASE logo or choose 'search -> Start a new search'), click the 'View more search options' link if it has not been previously clicked:



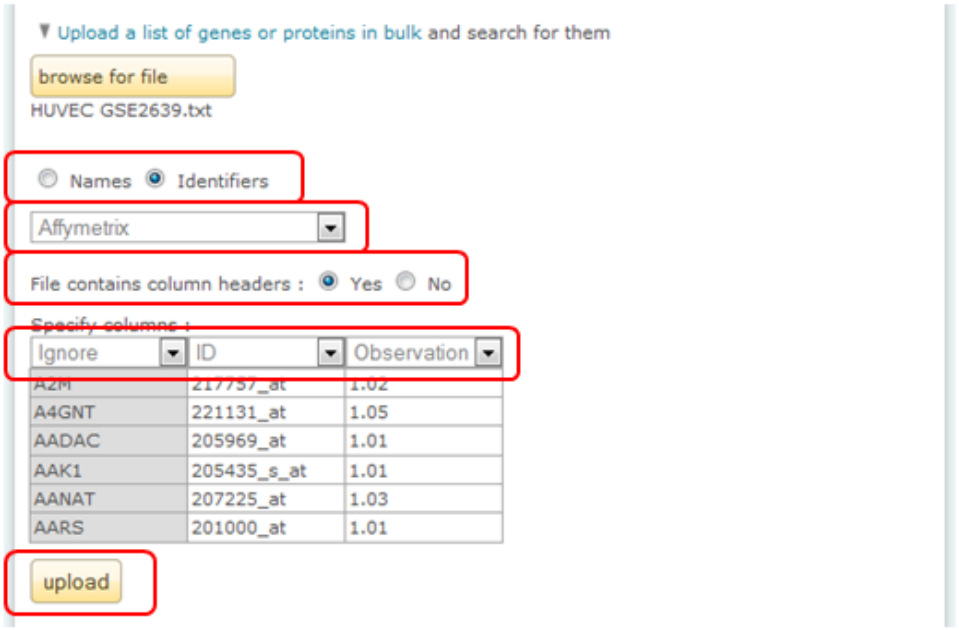
Select the 'Genes and proteins' radio button and then click the 'Upload a list of genes or proteins in bulk' link:



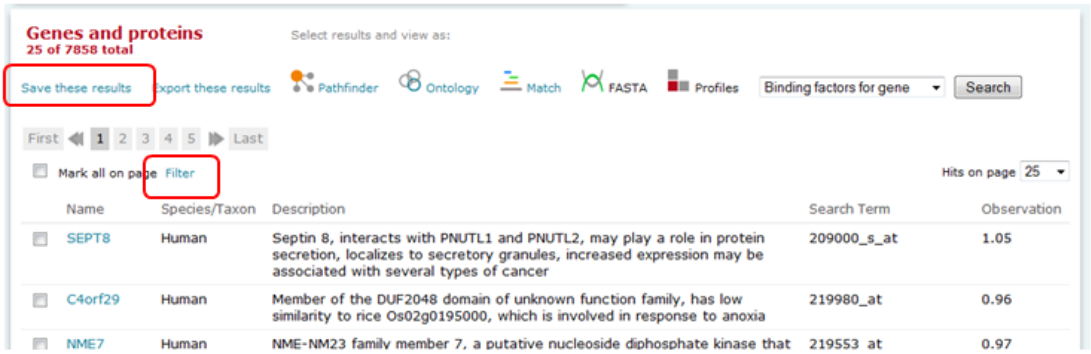
Click the yellow 'browse for file' button to select a tab-delimited or Excel file of gene names or identifiers:



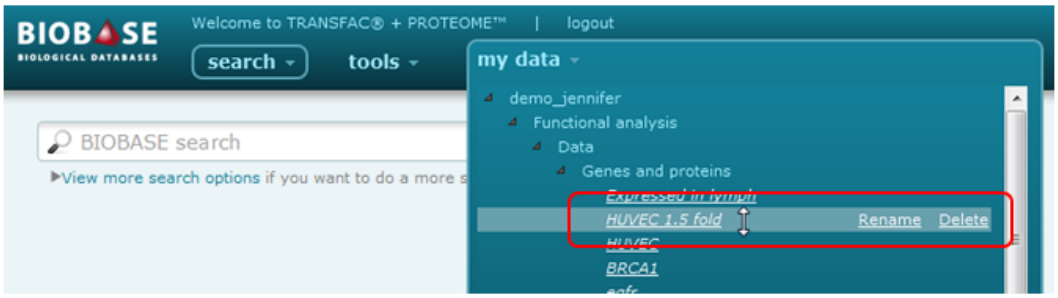
Once you have selected the file, use the prompts to specify whether your file contains a list of gene names or identifiers, the species (in the case of gene names) or the identifier type (in the case of identifiers), and whether your file contains a header column (and the first row should therefore be ignored). Using the preview provided of the file contents, you must specify which column contains the name or identifier to be used for matching (specify this column as 'ID' from the pull-down menu) and may optionally specify one column containing a numeric value such as fold-change (specify this column as 'Observation' from the pull-down menu). All other columns must be set to 'Ignore'. Finally, click the 'upload' button to upload the contents of the file.



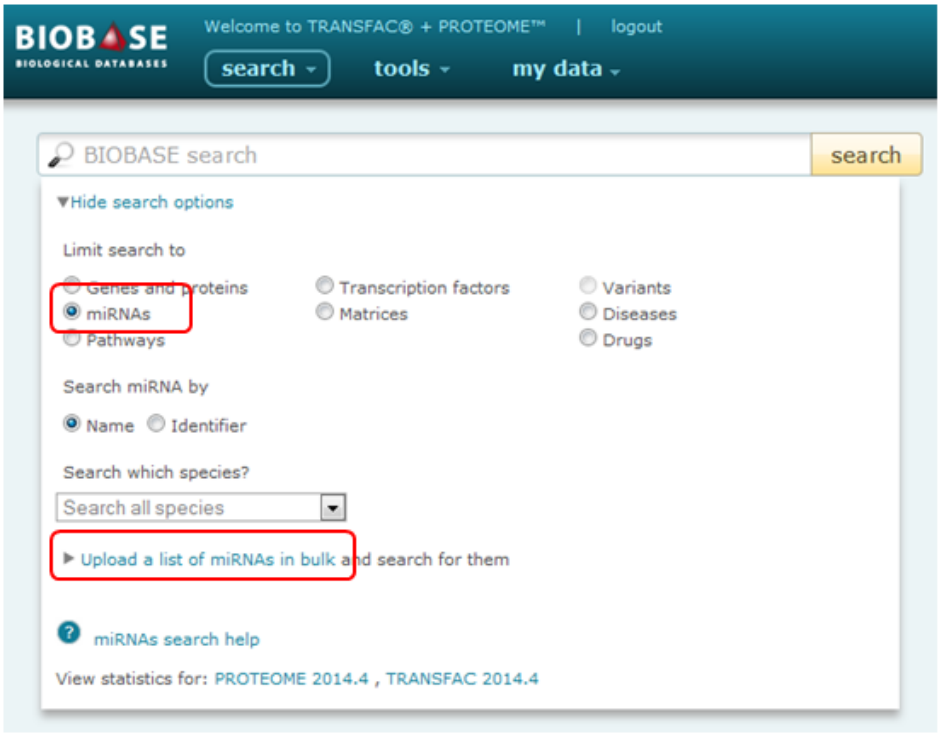
You will receive a message noting which values could not be matched to an entry in the database, and then the matched entries will be returned in the results section below. Once the results are returned, click the 'Save these results' link to save the list of genes for further analysis.


Genes and proteins
25 of 7858 total
Select results and view as:
Save these results
Export these results
Pathfinder
Ontology
Match
FASTA
Profiles
Binding factors for gene
Search
First 1 2 3 4 5 Last
Mark all on page
Filter
Hits on page 25
Name Species/Taxon Description Search Term Observation
SEPT8 Human Septin 8, interacts with PNUTL1 and PNUTL2, may play a role in protein secretion, localizes to secretory granules, increased expression may be associated with several types of cancer 209000_s_at 1.05
C4orf29 Human Member of the DUF2048 domain of unknown function family, has low similarity to rice Os02g0195000, which is involved in response to anoxia 219980_at 0.96
NME7 Human NME-NM23 family member 7, a putative nucleoside diphosphate kinase that 219553_at 0.97

To save a subset of the genes, use a combination of clicking column headers for sorting and the Filter link for filtering, then select the desired subset to be saved using the check boxes next to each entry. After saving, the data set can be accessed from the 'my data' menu.

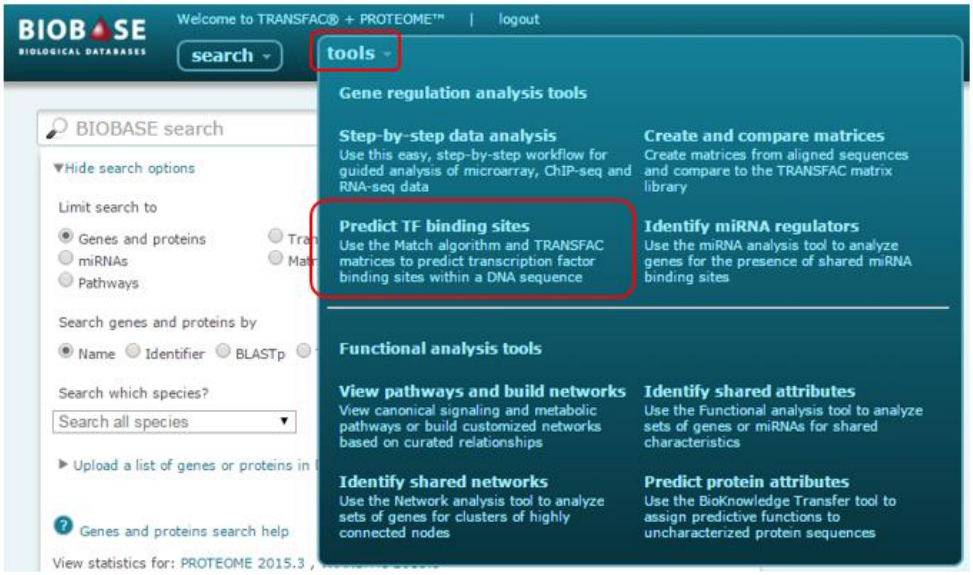


Note that unlike ExPlain, TRANSFAC supports uploading of mature miRNAs by name or identifier. To upload a list of mature miRNAs such as hsa-miR-155-5p (as opposed to a list of miRNA genes such as human MIR155), select the 'miRNAs' radio button, click the 'Upload a list of miRNAs in bulk' link and then proceed as described above for uploading a list of genes.

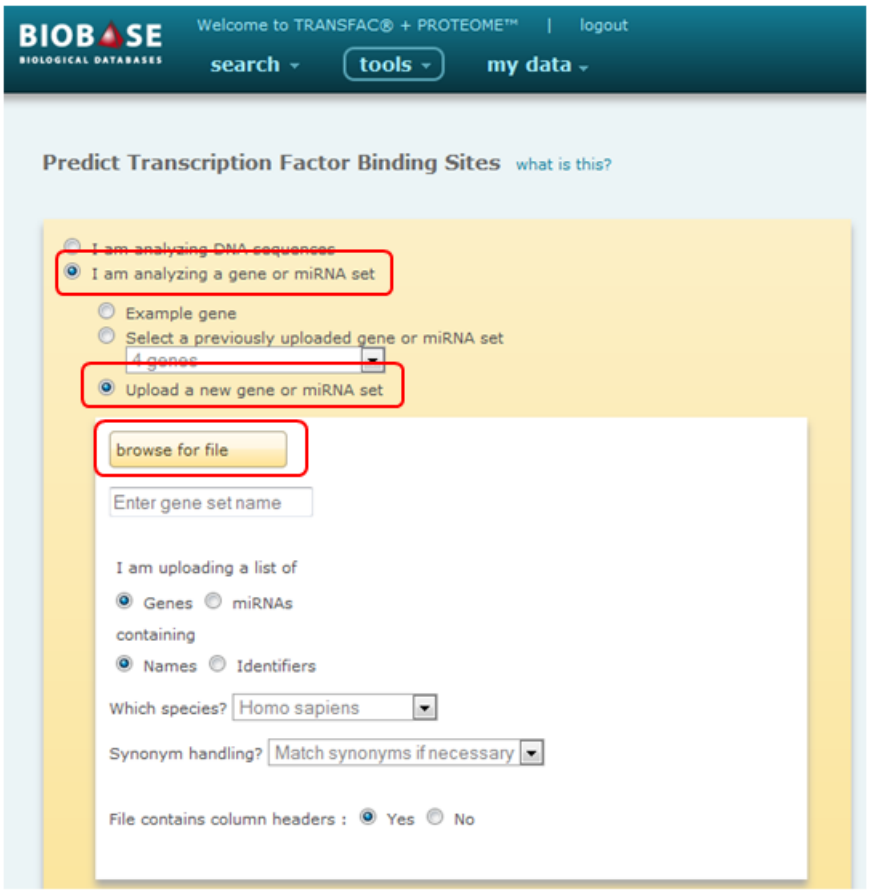


2. Match gene or miRNA set upload option

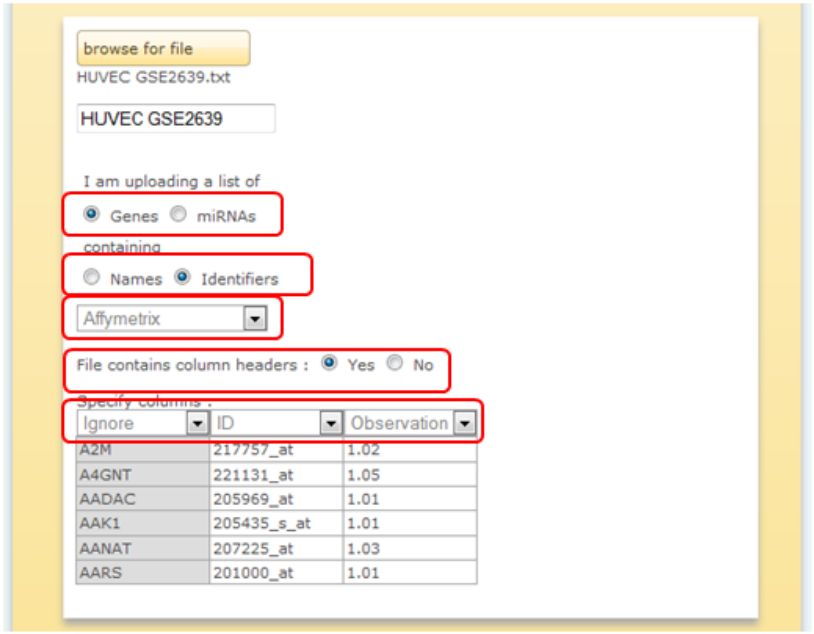
To upload a list of genes or miRNAs and launch a Match analysis in one step, from any location, choose 'tools -> Predict TF binding sites'.


BIOBASE
BIOLOGICAL DATABASES
Welcome to TRANSFAC® + PROTEOME™ | logout
search
tools
BIOBASE search
Hide search options
Limit search to
Genes and proteins
miRNAs
Pathways
Search genes and proteins by
Name
Identifier
BLASTp
Search which species?
Search all species
Upload a list of genes or proteins in
Genes and proteins search help
View statistics for: PROTEOME 2015.3
Gene regulation analysis tools
Step-by-step data analysis
Use this easy, step-by-step workflow for guided analysis of microarray, ChIP-seq and RNA-seq data
Create and compare matrices
Create matrices from aligned sequences and compare to the TRANSFAC matrix library
Predict TF binding sites
Use the Match algorithm and TRANSFAC matrices to predict transcription factor binding sites within a DNA sequence
Identify miRNA regulators
Use the miRNA analysis tool to analyze genes for the presence of shared miRNA binding sites
Functional analysis tools
View pathways and build networks
View canonical signaling and metabolic pathways or build customized networks based on curated relationships
Identify shared attributes
Use the Functional analysis tool to analyze sets of genes or miRNAs for shared characteristics
Identify shared networks
Use the Network analysis tool to analyze sets of genes for clusters of highly connected nodes
Predict protein attributes
Use the BioKnowledge Transfer tool to assign predictive functions to uncharacterized protein sequences

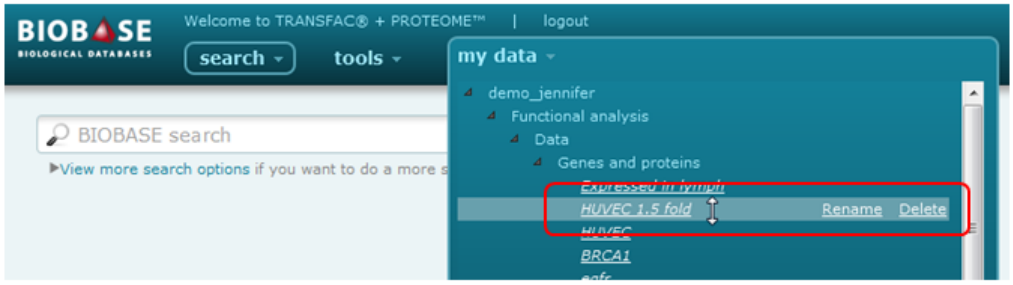
When the Match tool loads select the 'I am uploading a gene or miRNA set' radio button, select the 'Upload a new gene or miRNA set' radio button and then click the yellow 'browse for file' button to select a tab-delimited or Excel file of gene names or identifiers:



Once you have selected the file, use the prompts to specify whether your file contains a list of genes or miRNAs, whether it is a list of names or identifiers, the species (in the case of gene/miRNA names) or the identifier type (in the case of identifiers), and whether your file contains a header column (and the first row should therefore be ignored). Using the preview provided of the file contents, you must specify which column contains the name or identifier to be used for matching (specify this column as 'ID' from the pull-down menu) and may optionally specify one column containing a numeric value such as fold-change (specify this column as 'Observation' from the pull-down menu). All other columns must be set to 'Ignore'.

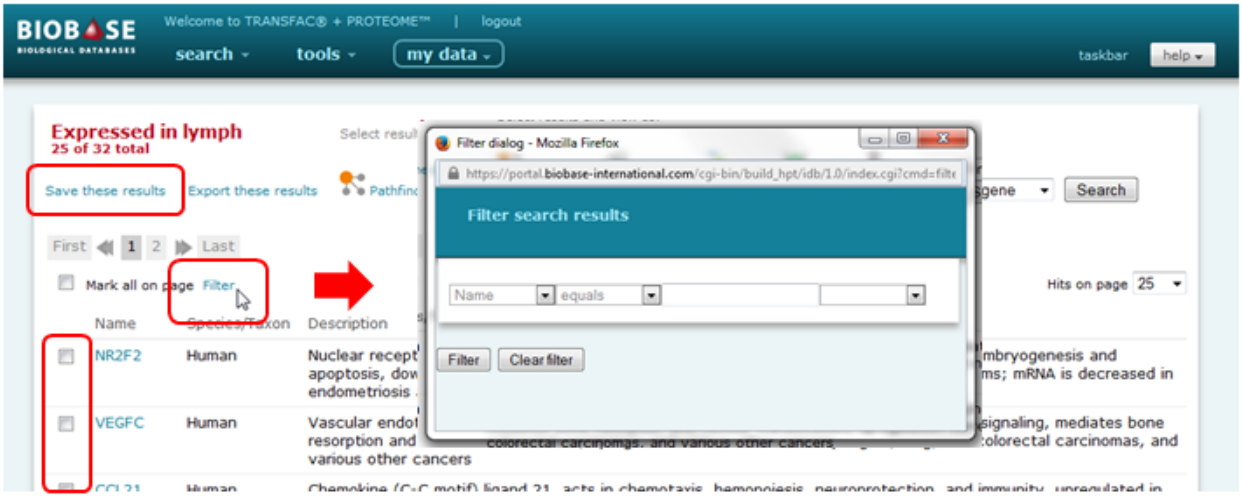


When you start the Match analysis, after filling in the required parameters, the data set will be automatically uploaded and saved to your user account in the 'my data' menu.



Creating data subsets

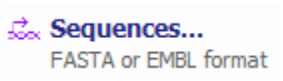
Whenever a list of Genes and proteins or miRNAs is displayed – whether that list is the direct result of a search or is a previously saved data set that has been loaded from the 'my data' menu – individual entries can be selected using the checkboxes, and then saved using the 'Save these results' link:

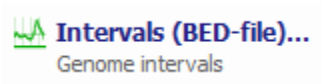


Alternatively the Filter link can be used to open a filter dialog box that allows you to actively filter the data set by any of the available columns, and then save the filtered data set.

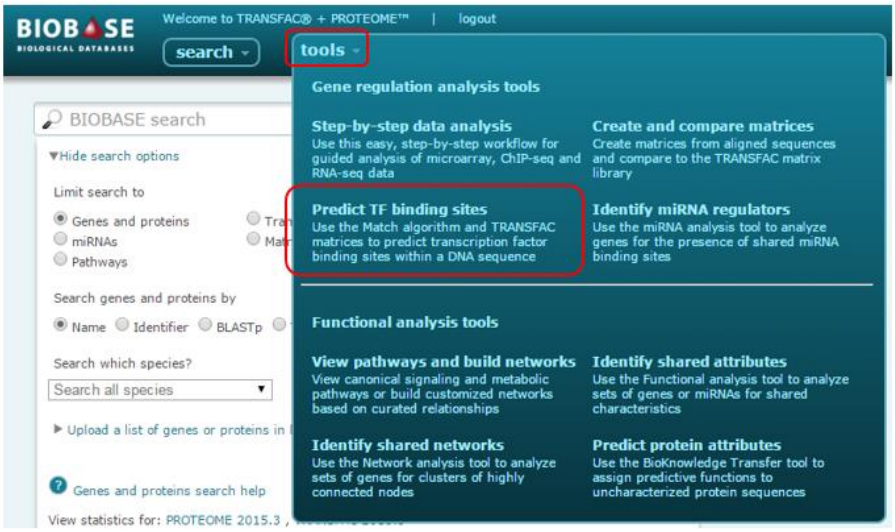
Loading sequences and intervals

The options described here replace the '**File -> Create new data -> Sequences**', '**File -> Load data from file -> Sequence**' and '**File -> Load data from file -> Intervals (BED-file)**' functions in ExPlain:

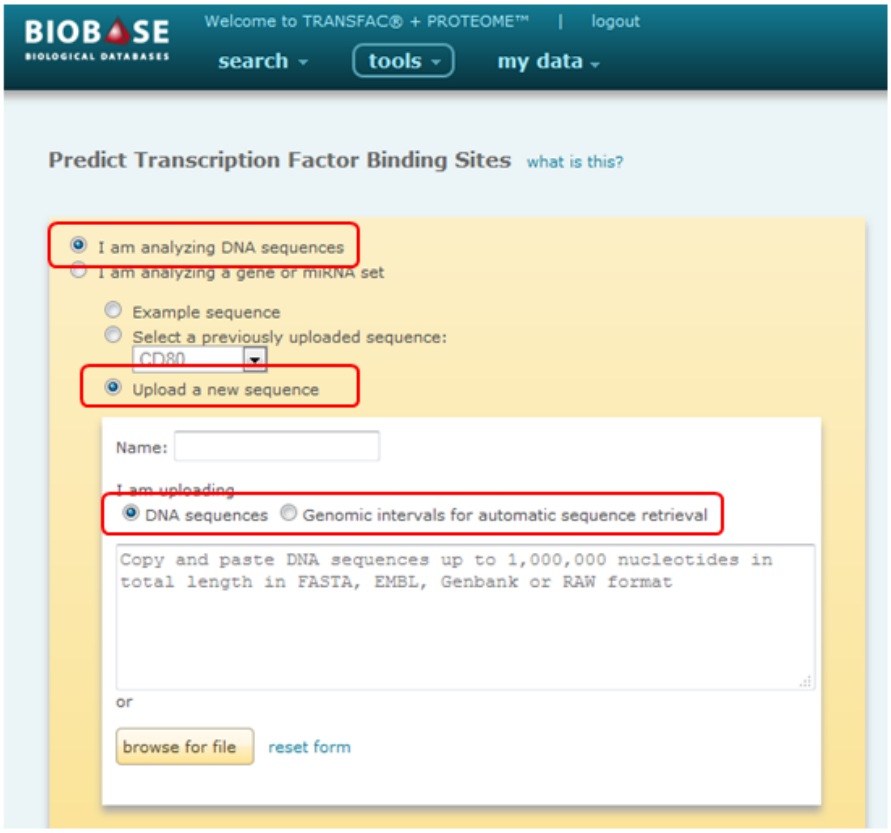




Raw sequences, as well as intervals for extracting sequences, are loaded through the Match tool. From any page within TRANSFAC, click the tools menu and then the Predict TF binding sites link:


BIOBASE
Welcome to TRANSFAC® + PROTEOME™ | logout
search
tools
Gene regulation analysis tools
Step-by-step data analysis
Use this easy, step-by-step workflow for guided analysis of microarray, ChIP-seq and RNA-seq data
Create and compare matrices
Create matrices from aligned sequences and compare to the TRANSFAC matrix library
Predict TF binding sites
Use the Match algorithm and TRANSFAC matrices to predict transcription factor binding sites within a DNA sequence
Identify miRNA regulators
Use the miRNA analysis tool to analyze genes for the presence of shared miRNA binding sites
Functional analysis tools
View pathways and build networks
View canonical signaling and metabolic pathways or build customized networks based on curated relationships
Identify shared attributes
Use the Functional analysis tool to analyze sets of genes or miRNAs for shared characteristics
Identify shared networks
Use the Network analysis tool to analyze sets of genes for clusters of highly connected nodes
Predict protein attributes
Use the BioKnowledge Transfer tool to assign predictive functions to uncharacterized protein sequences
BIOBASE search
Hide search options
Limit search to
Genes and proteins
miRNAs
Pathways
Search genes and proteins by
Name
Identifier
BLASTp
Search which species?
Search all species
Upload a list of genes or proteins in
Genes and proteins search help
View statistics for: PROTEOME 2015.3

Within the Match tool, select the 'I am analyzing DNA sequences' radio button and the 'Upload a new sequence' radio button. Then select whether you wish to upload 'DNA sequences' directly or to submit 'Genomic intervals for automatic sequence retrieval':

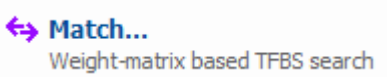


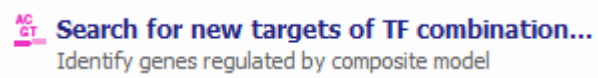
Any sequence or set of sequences up to 1,000,000 nucleotides in length can be uploaded in FASTA, EMBL, Genbank or RAW format. For human, mouse and rat species you can alternatively upload a list of

genomic coordinates in .bed format. The pull-down menu that appears when this option is selected will specify what genome version is supported. As of December 2014, the supported versions are human hg38/GRCh38, mouse mm10/GRCm38, and rat rn5/RGSC 5.0.

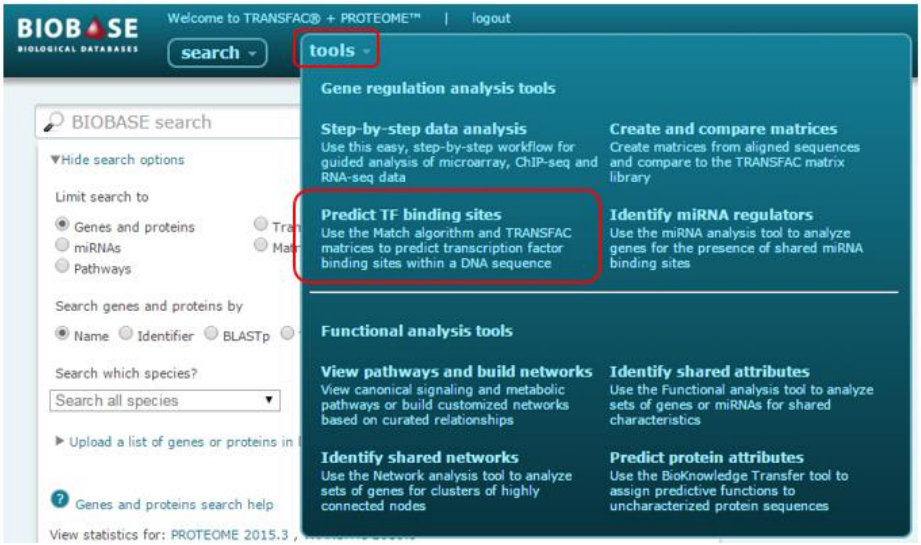
Match and Composite model search

TRANSFAC's Predict TF binding sites tool which includes the Match, Composite model and FMatch analysis options replaces the '**Analyze -> Binding sites search -> Match**' and '**Analyze -> Binding sites search -> Search for new targets of TF combination**' functions in ExPlain:

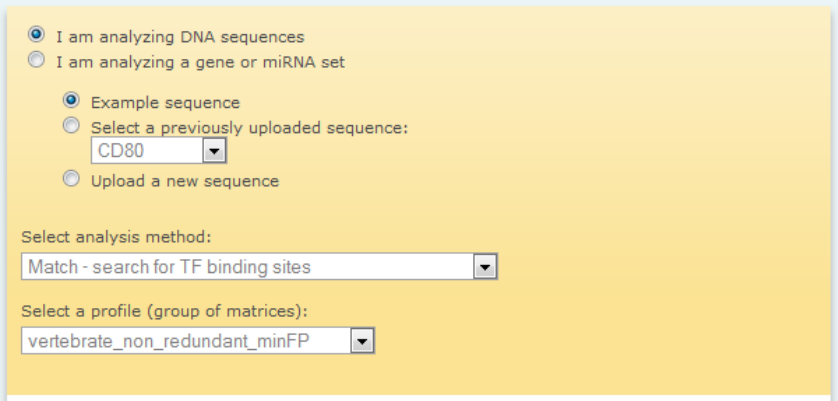




To access the Match tool, from any page within TRANSFAC, click the tools menu and then the Predict TF binding sites link:


BIOBASE
Welcome to TRANSFAC® + PROTEOME™ | logout
search
tools
Gene regulation analysis tools
Step-by-step data analysis
Use this easy, step-by-step workflow for guided analysis of microarray, ChIP-seq and RNA-seq data
Create and compare matrices
Create matrices from aligned sequences and compare to the TRANSFAC matrix library
Predict TF binding sites
Use the Match algorithm and TRANSFAC matrices to predict transcription factor binding sites within a DNA sequence
Identify miRNA regulators
Use the miRNA analysis tool to analyze genes for the presence of shared miRNA binding sites
Functional analysis tools
View pathways and build networks
View canonical signaling and metabolic pathways or build customized networks based on curated relationships
Identify shared attributes
Use the Functional analysis tool to analyze sets of genes or miRNAs for shared characteristics
Identify shared networks
Use the Network analysis tool to analyze sets of genes for clusters of highly connected nodes
Predict protein attributes
Use the BioKnowledge Transfer tool to assign predictive functions to uncharacterized protein sequences
BIOBASE search
Hide search options
Limit search to
Genes and proteins
miRNAs
Pathways
Search genes and proteins by
Name
Identifier
BLASTp
Search which species?
Search all species
Upload a list of genes or proteins in
Genes and proteins search help
View statistics for: PROTEOME 2015.3

There are three types of input that are required in the yellow section of the Match tool:

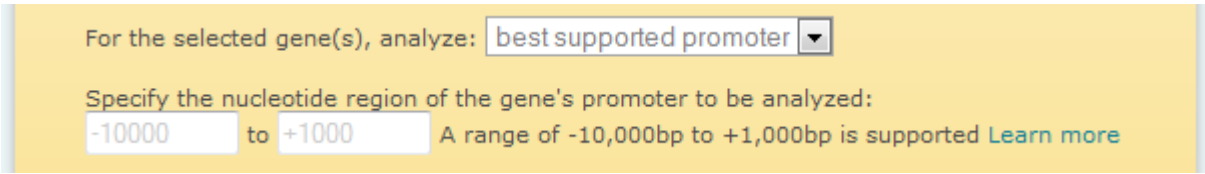

I am analyzing DNA sequences
I am analyzing a gene or miRNA set
Example sequence
Select a previously uploaded sequence:
CD80
Upload a new sequence
Select analysis method:
Match - search for TF binding sites
Select a profile (group of matrices):
vertebrate_non_redundant_minFP

1. The data set to be analyzed.

The tool provides the option to analyze DNA sequences or to analyze a gene or miRNA set. When 'I am analyzing DNA sequences' is selected, you are given the option to run Match against an example sequence, against a previously uploaded sequence or set of sequences or to upload a new sequence or set of sequences. Uploading sequences, either as raw sequences (i.e. FASTA, Genbank) or as genomic intervals, is described in Loading sequences and intervals.

When 'I am analyzing a gene or miRNA set' is selected, you are given the option to run Match against an example gene, against a previously uploaded list of genes or miRNAs or to upload a new list of genes or miRNAs. Despite the different entry point, the process for uploading a list of genes and miRNAs is the same as is described in Loading gene and miRNA sets.

When you choose the option to analyze a gene or miRNA set, Match will extract the promoter sequence(s) associated with the gene(s) or miRNA(s) from the underlying database and use those sequences for its analysis. You are provided with 2 additional parameters so that you can specify (1) whether you want Match to consider all available promoters for a gene or only the best supported promoter (which is defined as the promoter with the most significant cluster of Ensembl transcription start sites) and (2) what portion of the promoter sequence you want to consider. Like in ExPlain, TRANSFAC promoters are 11,000 bp in length, spanning from -10,000 base pairs upstream of the TSS to +1,000 base pairs downstream of the TSS. A typical range is -500 to +100.


For the selected gene(s), analyze:
best supported promoter
Specify the nucleotide region of the gene's promoter to be analyzed:
-10000
to
+1000
A range of -10,000bp to +1,000bp is supported
Learn more

2. The analysis method to be used.

Match - By default, the analysis method will be set to 'Match – search for TF binding sites'. When this method is selected, the Match algorithm will use the positional weight matrices in the selected profile to search your sequences for binding sites which meet the cut-off criteria. This option is equivalent to running 'Analyze -> Binding sites search -> Match' in ExPlain with no background set selected.

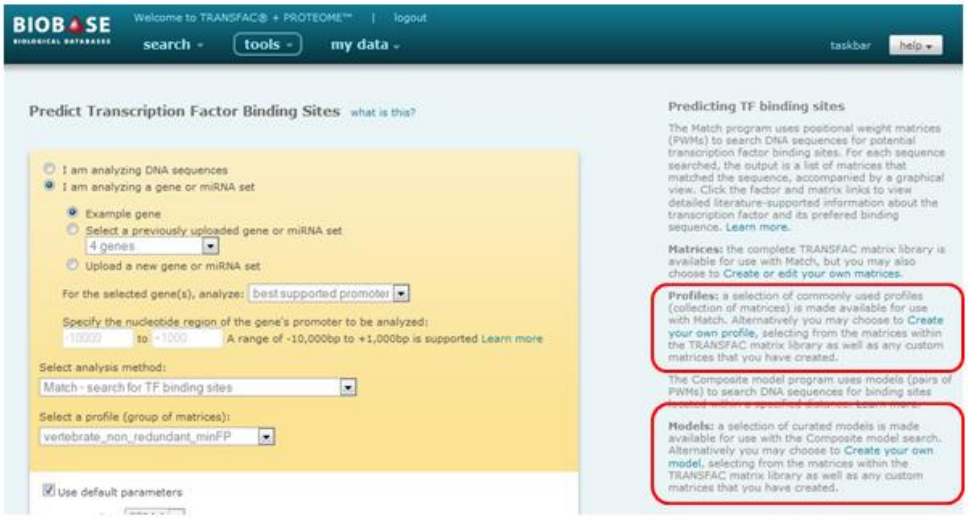
Composite model – When 'Composite model – search by pairs of TFs' is selected, the composite model algorithm (which is based on the Match algorithm), use models composed of pairs of matrices separated by a specified gap to search your sequences for co-occurring binding sites. This option is equivalent to running 'Analyze ->Binding sites search -> Search for new targets of TF combination' in ExPlain.

FMatch – When 'FMatch – search for over-represented binding sites' is selected, the Match algorithm will use the positional weight matrices in the selected profile to search your sequences as well as a set of background sequences for binding sites which meet the cut-off criteria and then reports those PWMs and sites which are over-represented in your sequence set compared to the background sequence set. This option is equivalent to running 'Analyze -> Binding sites search -> Match' in ExPlain with a background set.

Note that when you select the FMatch option you will be asked to additionally provide information about the background set to be used.

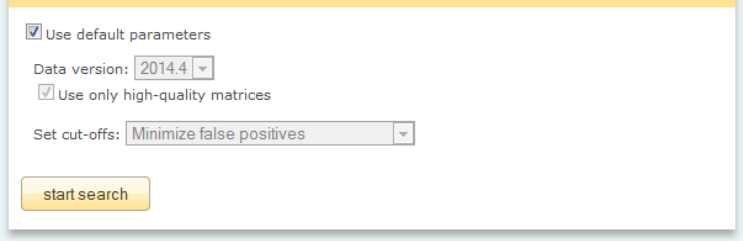
3. The profile (group of positional weight matrices) to be used.

As in ExPlain, a number of prepared profiles are provided, including the default Vertebrate Non-Redundant profile along with a number of tissue specific profiles. You can also create your own custom profiles from the collection of matrices within TRANSFAC or from your own uploaded matrices using the Profile creation tool accessed through the help section on the right-hand-side of the screen:


BIOBASE
Welcome to TRANSFAC® + PROTEOME™ | logout
search
tools
my data
taskbar
help
Predict Transcription Factor Binding Sites what is this?
I am analyzing DNA sequences
I am analyzing a gene or miRNA set
Example gene
Select a previously uploaded gene or miRNA set
4 genes
Upload a new gene or miRNA set
For the selected gene(s), analyze: best supported promoter
Specify the nucleotide region of the gene's promoter to be analyzed:
-10000 to +1000 A range of -10,000bp to +1,000bp is supported Learn more
Select analysis method:
Match - search for TF binding sites
Select a profile (group of matrices):
vertebrate_non_redundant_minFP
Use default parameters
Predicting TF binding sites
The Match program uses positional weight matrices (PWMs) to search DNA sequences for potential transcription factor binding sites. For each sequence searched, the output is a list of matrices that matched the sequence, accompanied by a graphical view. Click the factor and matrix links to view detailed literature-supported information about the transcription factor and its prefered binding sequence. Learn more.
Matrices: the complete TRANSFAC matrix library is available for use with Match, but you may also choose to Create or edit your own matrices.
Profiles: a selection of commonly used profiles (collection of matrices) is made available for use with Match. Alternatively you may choose to Create your own profile, selecting from the matrices within the TRANSFAC matrix library as well as any custom matrices that you have created.
The Composite model program uses models (pairs of PWMs) to search DNA sequences for binding sites
Models: a selection of curated models is made available for use with the Composite model search. Alternatively you may choose to Create your own model, selecting from the matrices within the TRANSFAC matrix library as well as any custom matrices that you have created.

Similarly, when using the Composite model option, a number of prepared models are provided or you can create your own models using the Model creation tool also accessed through the help section on the right-hand-side of the screen.

Once the required inputs have been specified, you can launch the analysis and it will be run with the default parameters which are set in the white section of the Match tool:



If you wish to make any changes to the parameters, de-select the 'Use default parameters' and make the desired changes. The parameter options differ based on the analysis method selected, but collectively cover the following:

1. Data version

   TRANSFAC supports the four most recent data versions. By default the current version is selected, but previous versions may be selected using the pull-down menu.

2. Use only high quality matrices

By default, this option is used to exclude matrices which generate particularly high numbers of false positives from Match and FMatch analyses. For more information about how high versus low quality matrices are defined, please see the 'BKL Search and Tools -> TF Binding Site Prediction -> Cut-off Values' page of the TRANSFAC user manual. De-selecting this option will allow the low quality matrices to be included in the analysis.

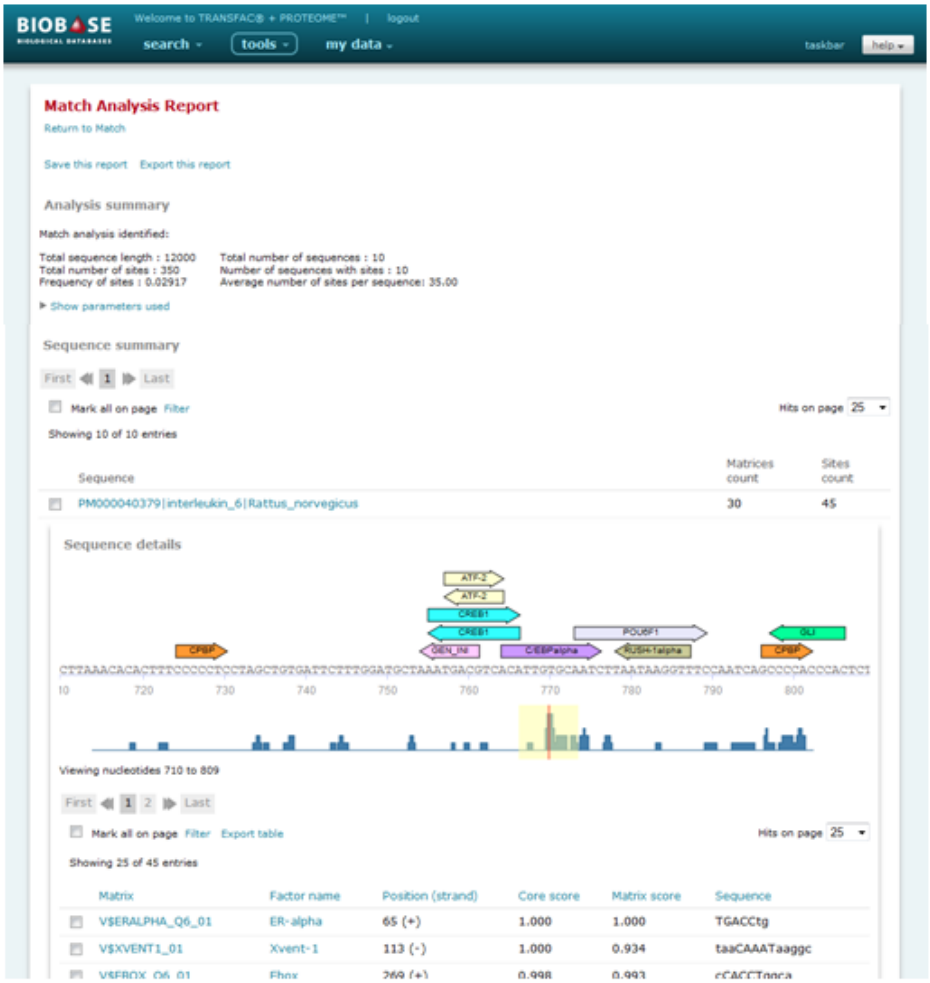
3. Set cut-offs

Depending on the profile selected (such as for the tissue specific profiles), you may have the option to modify the cut-off settings. When this option is enabled, you are able to select whether you want to use the minFP, minFN or minSUM cut-offs, to use the cut-offs from the profile (which for prepared profiles are usually the minFN cut-offs), or to manually set your own cut-offs.

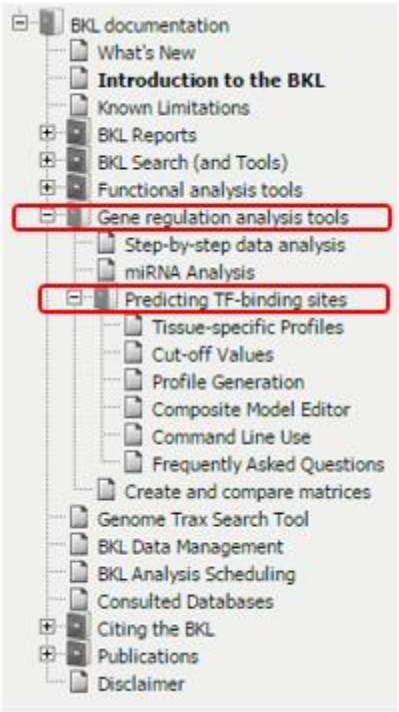
4. p-value threshold

As FMatch analyzes two sequence or gene sets in comparison, a p-value is calculated for over-representation of sites for particular matrices in the analyzed set versus the background set. FMatch compares the Match result of the two sets and optimizes the cut-offs for each matrix used in the analysis to reach the best separation between the two sets. Only those matrices for which the p-value for over-representation of the sites in the experimental set fits the p-value threshold are reported. By default the p-value is set to 0.01. To make the p-value more or less stringent, type the new p-value into the box.

When you are ready, launch the analysis by clicking the 'start search' button. Your analysis will be forwarded to the taskbar. If you keep the taskbar window open until the analysis completes, the analysis results will automatically be loaded within the open window.

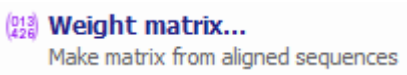


For a detailed description of the analysis report, please see the ‘BKL Search and Tools -> Predicting TF-binding sites’ page of the TRANSFAC user manual.

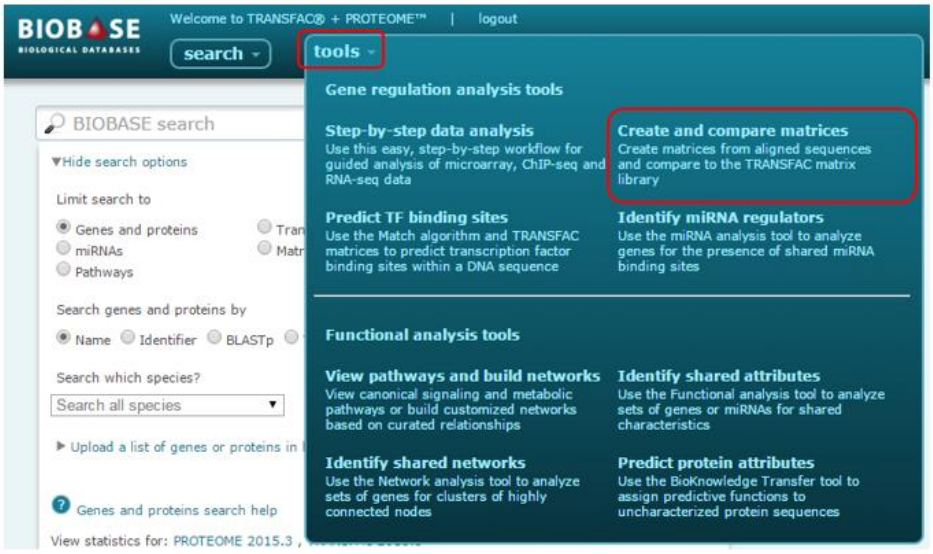


Loading matrices

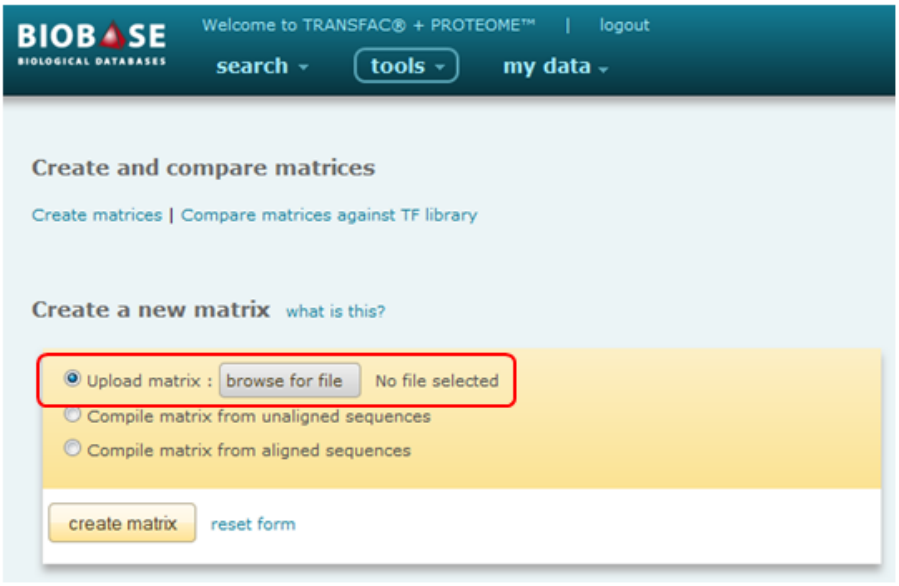
TRANSFAC's Create and compare matrices tool replaces the '**File -> Create new data -> Weight matrix**' option in ExPlain:



To load a custom matrix in TRANSFAC .dat format, or to create a matrix from a set of aligned or unaligned sequences, use the Create and compare matrices tools which is accessed from any page within TRANSFAC by clicking the tools menu followed by clicking the 'Create and compare matrices' link:


BIOBASE
BIOLOGICAL DATABASES
Welcome to TRANSFAC® + PROTEOME™ | logout
search
tools
Gene regulation analysis tools
Step-by-step data analysis
Use this easy, step-by-step workflow for guided analysis of microarray, ChIP-seq and RNA-seq data
Create and compare matrices
Create matrices from aligned sequences and compare to the TRANSFAC matrix library
Predict TF binding sites
Use the Match algorithm and TRANSFAC matrices to predict transcription factor binding sites within a DNA sequence
Identify miRNA regulators
Use the miRNA analysis tool to analyze genes for the presence of shared miRNA binding sites
Functional analysis tools
View pathways and build networks
View canonical signaling and metabolic pathways or build customized networks based on curated relationships
Identify shared attributes
Use the Functional analysis tool to analyze sets of genes or miRNAs for shared characteristics
Identify shared networks
Use the Network analysis tool to analyze sets of genes for clusters of highly connected nodes
Predict protein attributes
Use the BioKnowledge Transfer tool to assign predictive functions to uncharacterized protein sequences
BIOBASE search
Hide search options
Limit search to
Genes and proteins
miRNAs
Pathways
Search genes and proteins by
Name
Identifier
BLASTp
Search which species?
Search all species
Upload a list of genes or proteins in
Genes and proteins search help
View statistics for: PROTEOME 2015.3

To upload a matrix in TRANSFAC .dat format, select the 'Upload matrix' radio button, click the 'browse for file' button, select the file to be uploaded and click the 'create matrix' button:



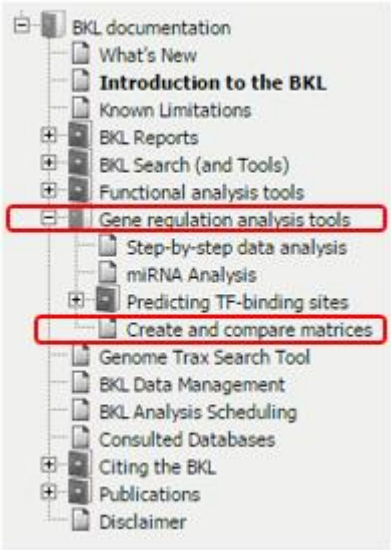
A nucleotide position frequency table and consensus sequence logo preview will be created and you will be prompted to click the 'save matrix and specify cut-off values' button to finalize the upload:

BIOBASE
BIOLOGICAL DATABASES
Welcome to TRANSFAC® + PROTEOME™ | logout
search
tools
my data
Create matrices
Matrix name : CREB_1105_03
Date created :
Identifier : V$CREB_1105_03
Nucleotide position frequency
A C G T Consensus
1 1 3 20 T
0 1 21 3 G
24 1 0 0 A
0 25 0 0 C
0 2 23 0 G
0 2 1 22 T
11 13 1 0 M
16 1 2 6 A
Save matrix and specify cut-off values
(Note that this operation may take a few minutes)

When the process is complete (this may take a few minutes due to the calculation of the cut-off values) you will be notified that the matrix has been saved to your library. It will now be listed in the 'Gene regulation analysis -> Data -> Matrices -> Uploaded matrices' folder of the 'my data' menu:

and will also be listed in the Profile creation tool (accessed from the Match tool page) which can be used to create a profile containing the matrix for use by Match.

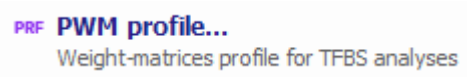
Alternatively you can create a matrix using a set of aligned or unaligned sequences. For a description of how to create a matrix from a set of aligned sequences, please see the 'Gene regulation analysis tools -> Create and compare matrices' page of the TRANSFAC user manual.



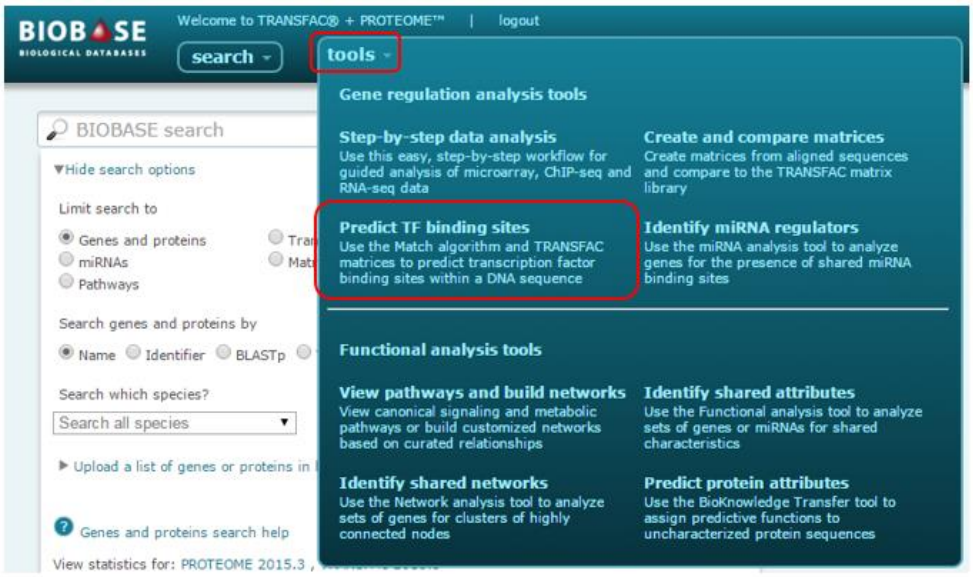
For a description of how to create a matrix from a set of unaligned sequences, please see Finding novel motifs (Seeder) in this document.

Creating profiles

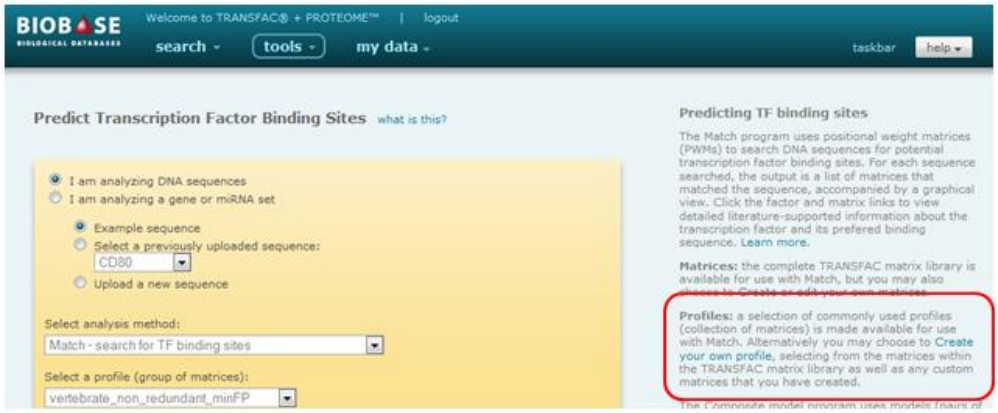
TRANSFAC's Profile creation tool replaces the '**File -> Create new data -> PWM profile**' option in ExPlain:



The tool for creating custom profiles is accessed through Match. From any page within TRANSFAC, click the tools menu and then the 'Predict TF binding sites' link:

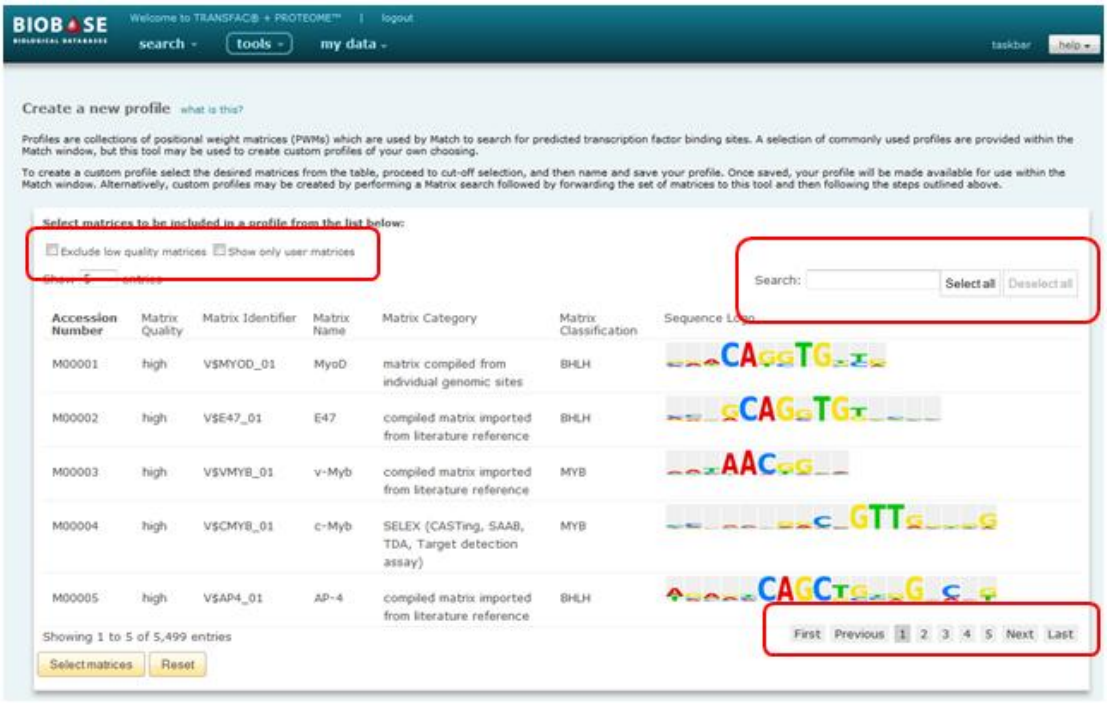

BIOBASE
Welcome to TRANSFAC® + PROTEOME™ | logout
search
tools
Gene regulation analysis tools
Step-by-step data analysis
Use this easy, step-by-step workflow for guided analysis of microarray, ChIP-seq and RNA-seq data
Create and compare matrices
Create matrices from aligned sequences and compare to the TRANSFAC matrix library
Predict TF binding sites
Use the Match algorithm and TRANSFAC matrices to predict transcription factor binding sites within a DNA sequence
Identify miRNA regulators
Use the miRNA analysis tool to analyze genes for the presence of shared miRNA binding sites
Functional analysis tools
View pathways and build networks
View canonical signaling and metabolic pathways or build customized networks based on curated relationships
Identify shared attributes
Use the Functional analysis tool to analyze sets of genes or miRNAs for shared characteristics
Identify shared networks
Use the Network analysis tool to analyze sets of genes for clusters of highly connected nodes
Predict protein attributes
Use the BioKnowledge Transfer tool to assign predictive functions to uncharacterized protein sequences
BIOBASE search
Hide search options
Limit search to
Genes and proteins
miRNAs
Pathways
Search genes and proteins by
Name
Identifier
BLASTp
Search which species?
Search all species
Upload a list of genes or proteins in
Genes and proteins search help
View statistics for: PROTEOME 2015.3

And then click the 'Create your own profile' link in the right-hand-side of the page:

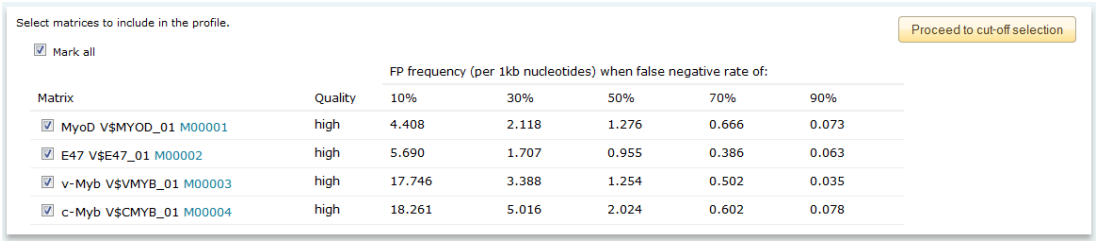

BIOBASE
Welcome to TRANSFAC® + PROTEOME™ | logout
search
tools
my data
taskbar
help
Predict Transcription Factor Binding Sites what is this?
I am analyzing DNA sequences
I am analyzing a gene or miRNA set
Example sequence
Select a previously uploaded sequence:
CD80
Upload a new sequence
Select analysis method:
Match - search for TF binding sites
Select a profile (group of matrices):
vertebrate_non_redundant_minFP
Predicting TF binding sites
The Match program uses positional weight matrices (PWMs) to search DNA sequences for potential transcription factor binding sites. For each sequence searched, the output is a list of matrices that matched the sequence, accompanied by a graphical view. Click the factor and matrix links to view detailed literature-supported information about the transcription factor and its prefered binding sequence. Learn more.
Matrices: the complete TRANSFAC matrix library is available for use with Match, but you may also
Profiles: a selection of commonly used profiles (collection of matrices) is made available for use with Match. Alternatively you may choose to Create your own profile, selecting from the matrices within the TRANSFAC matrix library as well as any custom matrices that you have created.

Note that you can also access the Profile creation tool via the Create and compare matrices tool.

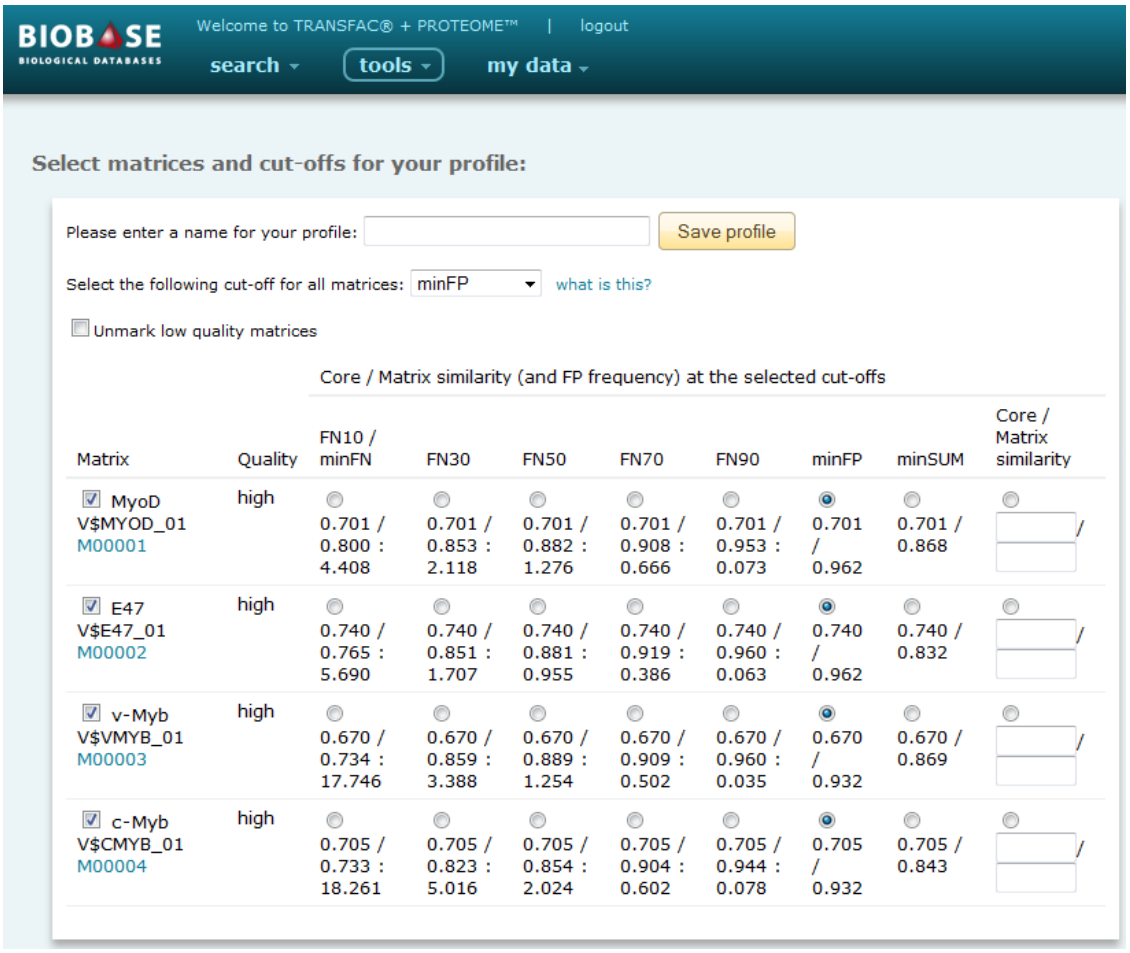
When the tool loads you can browse the library of matrices by scrolling through the individual pages, but you can more easily narrow the list by searching for specific matrices by name or by filtering the list to Exclude low quality matrices or to Show only user matrices (matrices that you have uploaded):


BIOBASE
search
tools
my data
help
Create a new profile
Exclude low quality matrices
Show only user matrices
Search:
Select all
Deselect all
Accession Number
Matrix Quality
Matrix Identifier
Matrix Name
Matrix Category
Matrix Classification
Sequence Logo
M00001 high V$MYOD_01 MyoD matrix compiled from individual genomic sites BHLH
M00002 high V$E47_01 E47 compiled matrix imported from literature reference BHLH
M00003 high V$VMYB_01 v-Myb compiled matrix imported from literature reference MYB
M00004 high V$CMYB_01 c-Myb SELEX (CASTing, SAAB, TDA, Target detection assay) MYB
M00005 high V$AP4_01 AP-4 compiled matrix imported from literature reference BHLH
Showing 1 to 5 of 5,499 entries
First Previous 1 2 3 4 5 Next Last
Select matrices
Reset

Once you have selected the matrix or matrices that you would like to include in your profile, click the 'Select matrices' button. You will be given a preview of the set that you have selected:

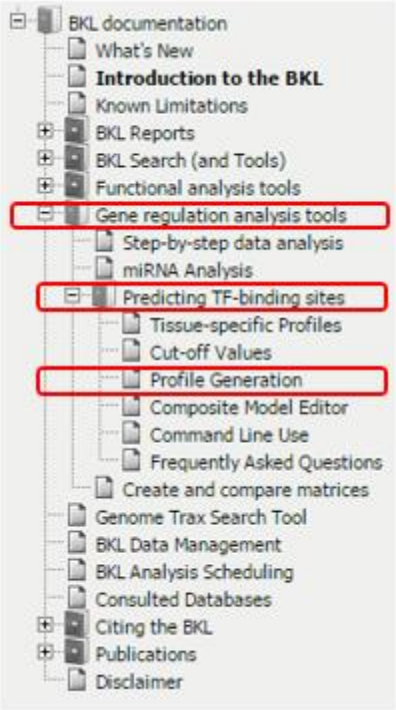

Select matrices to include in the profile.
Proceed to cut-off selection
Mark all
FP frequency (per 1kb nucleotides) when false negative rate of:
Matrix Quality 10% 30% 50% 70% 90%
MyoD V$MYOD_01 M00001 high 4.408 2.118 1.276 0.666 0.073
E47 V$E47_01 M00002 high 5.690 1.707 0.955 0.386 0.063
v-Myb V$VMYB_01 M00003 high 17.746 3.388 1.254 0.502 0.035
c-Myb V$CMYB_01 M00004 high 18.261 5.016 2.024 0.602 0.078

You can continue to add to or edit the set. When you are satisfied with the set, click the 'Proceed to cut-off selection' button. On the next screen you will be asked to name your profile and to select the cut-off values to be used:


BIOBASE
BIOLOGICAL DATABASES
Welcome to TRANSFAC® + PROTEOME™ | logout
search
tools
my data
Select matrices and cut-offs for your profile:
Please enter a name for your profile:
Save profile
Select the following cut-off for all matrices: minFP
what is this?
Unmark low quality matrices
Core / Matrix similarity (and FP frequency) at the selected cut-offs
Matrix
Quality
FN10 / minFN
FN30
FN50
FN70
FN90
minFP
minSUM
Core / Matrix similarity
MyoD V$MYOD_01 M00001 high 0.701 / 0.800 : 4.408 0.701 / 0.853 : 2.118 0.701 / 0.882 : 1.276 0.701 / 0.908 : 0.666 0.701 / 0.953 : 0.073 0.701 / 0.962 0.701 / 0.868
E47 V$E47_01 M00002 high 0.740 / 0.765 : 5.690 0.740 / 0.851 : 1.707 0.740 / 0.881 : 0.955 0.740 / 0.919 : 0.386 0.740 / 0.960 : 0.063 0.740 / 0.962 0.740 / 0.832
v-Myb V$VMYB_01 M00003 high 0.670 / 0.734 : 17.746 0.670 / 0.859 : 3.388 0.670 / 0.889 : 1.254 0.670 / 0.909 : 0.502 0.670 / 0.960 : 0.035 0.670 / 0.932 0.670 / 0.869
c-Myb V$CMYB_01 M00004 high 0.705 / 0.733 : 18.261 0.705 / 0.823 : 5.016 0.705 / 0.854 : 2.024 0.705 / 0.904 : 0.602 0.705 / 0.944 : 0.078 0.705 / 0.932 0.705 / 0.843

When you are finished, click the ‘Save profile’ button. The profile will now appear in the profile pull-down menu of the Match tool.

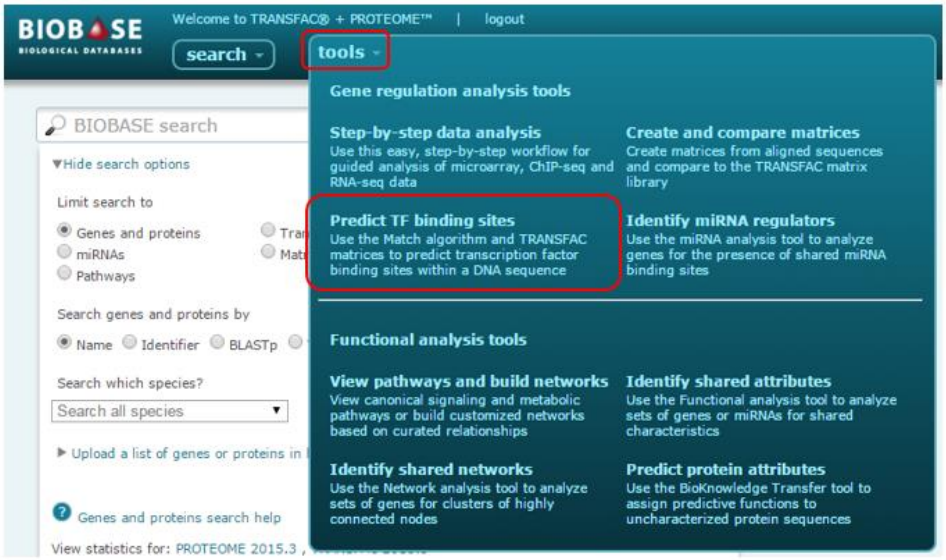
For a more detailed description of how to create a profile, please see the ‘Gene regulation analysis tools -> Predict TF-binding sites -> Profile Generation’ page of the TRANSFAC user manual.



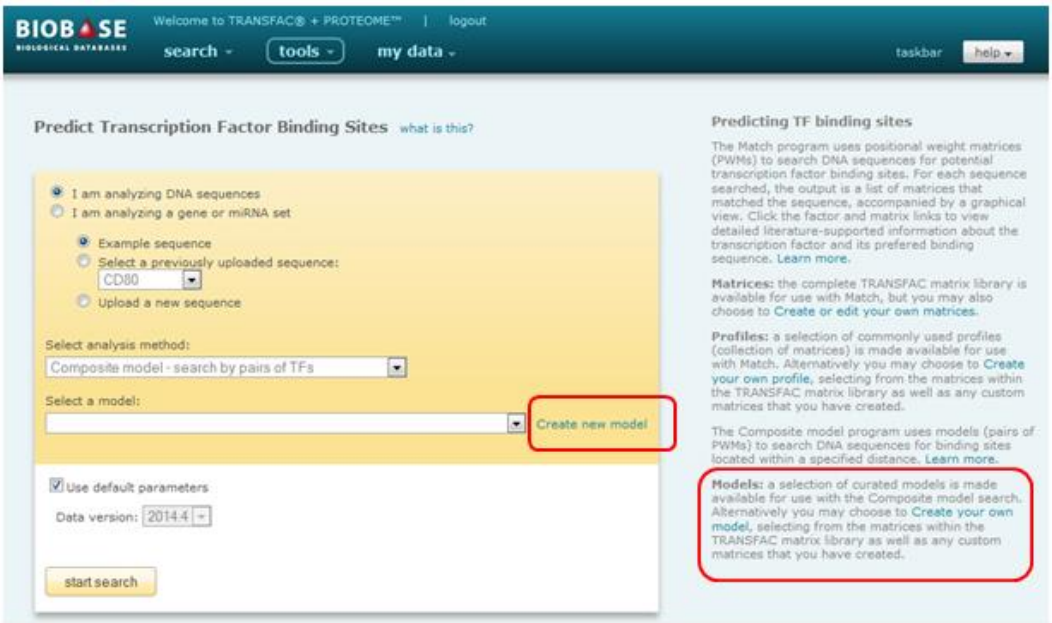
Creating composite models

TRANSFAC's Model creation tool replaces the '**File -> Create new data -> Composite model**' option in ExPlain:

The tool for creating custom models is accessed through Match. From any page within TRANSFAC, click the tools menu and then the 'Predict TF binding sites' link:

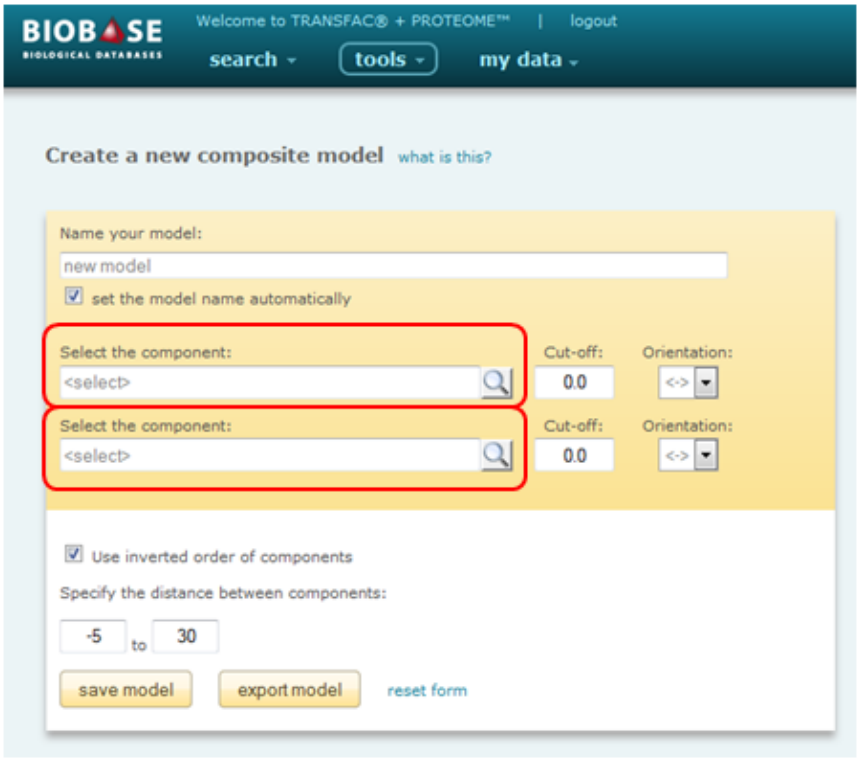

BIOBASE
Welcome to TRANSFAC® + PROTEOME™ | logout
search
tools
Gene regulation analysis tools
Step-by-step data analysis
Use this easy, step-by-step workflow for guided analysis of microarray, ChIP-seq and RNA-seq data
Create and compare matrices
Create matrices from aligned sequences and compare to the TRANSFAC matrix library
Predict TF binding sites
Use the Match algorithm and TRANSFAC matrices to predict transcription factor binding sites within a DNA sequence
Identify miRNA regulators
Use the miRNA analysis tool to analyze genes for the presence of shared miRNA binding sites
Functional analysis tools
View pathways and build networks
View canonical signaling and metabolic pathways or build customized networks based on curated relationships
Identify shared attributes
Use the Functional analysis tool to analyze sets of genes or miRNAs for shared characteristics
Identify shared networks
Use the Network analysis tool to analyze sets of genes for clusters of highly connected nodes
Predict protein attributes
Use the BioKnowledge Transfer tool to assign predictive functions to uncharacterized protein sequences
BIOBASE search
Hide search options
Limit search to
Genes and proteins
miRNAs
Pathways
Search genes and proteins by
Name
Identifier
BLASTp
Search which species?
Search all species
Upload a list of genes or proteins in
Genes and proteins search help
View statistics for: PROTEOME 2015.3

And then click the 'Create your own model link' in the right-hand-side of the page or the 'Create new model' link next to the model pull-down menu that appears when 'Composite model – search by pairs of TFs' is selected as the analysis method:


BIOBASE
Welcome to TRANSFAC® + PROTEOME™ | logout
search
tools
my data
taskbar
help
Predict Transcription Factor Binding Sites what is this?
I am analyzing DNA sequences
I am analyzing a gene or miRNA set
Example sequence
Select a previously uploaded sequence:
CD80
Upload a new sequence
Select analysis method:
Composite model - search by pairs of TFs
Select a model:
Create new model
Use default parameters
Data version: 2014.4
start search
Predicting TF binding sites
The Match program uses positional weight matrices (PWMs) to search DNA sequences for potential transcription factor binding sites. For each sequence searched, the output is a list of matrices that matched the sequence, accompanied by a graphical view. Click the factor and matrix links to view detailed literature-supported information about the transcription factor and its prefered binding sequence. Learn more.
Matrices: the complete TRANSFAC matrix library is available for use with Match, but you may also choose to Create or edit your own matrices.
Profiles: a selection of commonly used profiles (collection of matrices) is made available for use with Match. Alternatively you may choose to Create your own profile, selecting from the matrices within the TRANSFAC matrix library as well as any custom matrices that you have created.
The Composite model program uses models (pairs of PWMs) to search DNA sequences for binding sites located within a specified distance. Learn more.
Models: a selection of curated models is made available for use with the Composite model search. Alternatively you may choose to Create your own model, selecting from the matrices within the TRANSFAC matrix library as well as any custom matrices that you have created.

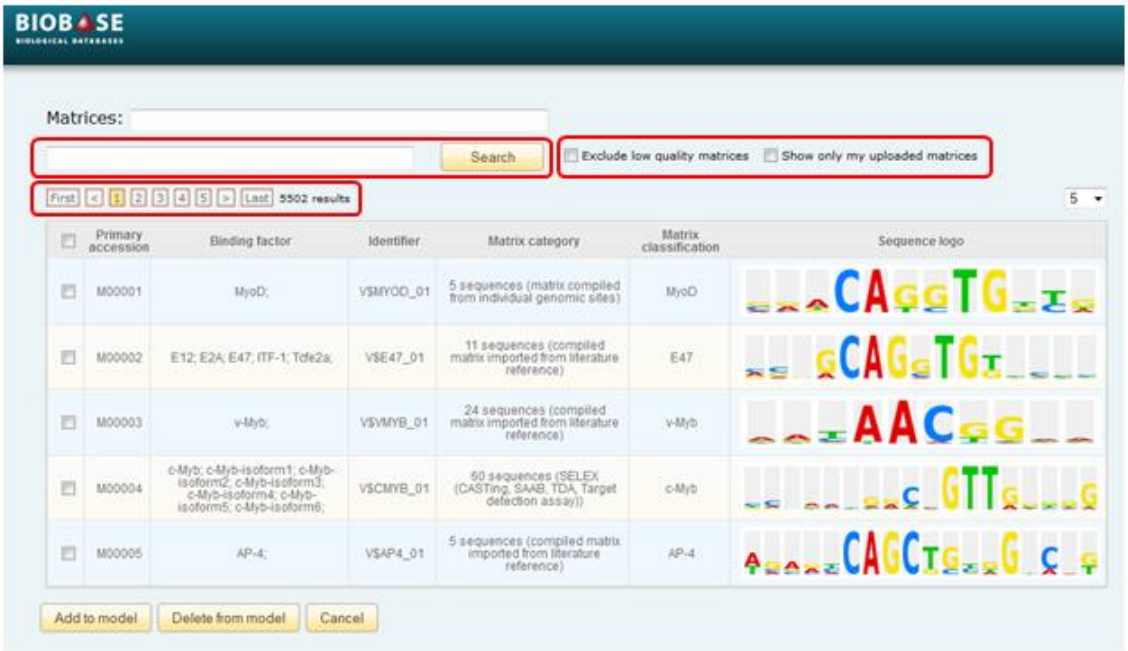
Note that you can also access the Profile creation tool via the Create and compare matrices tool.

When the tool loads you can select the pair of components (matrices) that you would like to include in your model:



When you click the magnifying glass icon for a component a matrix selection window will open.

When the window opens you can browse the library of matrices by scrolling through the individual pages, but you can more easily narrow the list by searching for specific matrices by name or by filtering the list to Exclude low quality matrices or to Show only user matrices (matrices that you have uploaded):



When you have selected the desired matrix or matrices (more than one matrix can be selected to represent a component), click the 'Add to model' button. The recommended minFN cut-off for the matrix or matrices is specified by default, but can be over-ridden by typing the desired cut-off into the Cut-off box. Directionless orientation is specified by default, but forward or reverse orientation can alternately be specified by selecting the desired option from the Orientation pull-down menu.

Repeat this process for the second component.

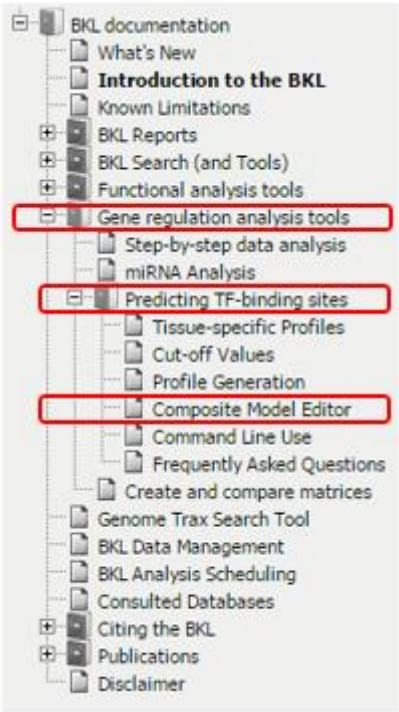
At this point you can save the model, keeping the remaining default parameters or by adapting them as desired:

1. Order of components – When the 'Use inverted order of components' parameter is checked (the default setting), the order in which the components are identified within the submitted sequence will not be considered when determining a match. For example, if Matrix A is selected as component 1 and Matrix B is selected as component 2, the model will be returned in the results regardless of whether the order is A -> B or B -> A, as long as the cut-off criteria are met. When the 'Use inverted order of components' selection is turned off, the order in which the components are identified will determine whether a match is made. For example, if Matrix A is selected as component 1 and Matrix B is selected as component 2, and 'Use inverted order of components' is turned off, only A -> B models will be returned in the results when the cut-off criteria is met. B -> A models will be ignored.

2. Distance between components - This parameter specifies the maximum distance, in nucleotides, that may exist between the two components in order for a pair of binding sites to be identified. A negative starting distance specifies the number of nucleotides by which the two matrices may overlap. A default distance of -5 to 30 is set, which is appropriate for most analyses.

Once the model is saved, it will appear in the 'my data' menu as well as the model pull-down menu of the Match tool when 'Composite model – search by pairs of TFs' is selected as the analysis method.

For a more detailed description of how to create a profile, please see the 'Gene regulation analysis tools-> Predict TF-binding sites -> Composite Model Editor' page of the TRANSFAC user manual.

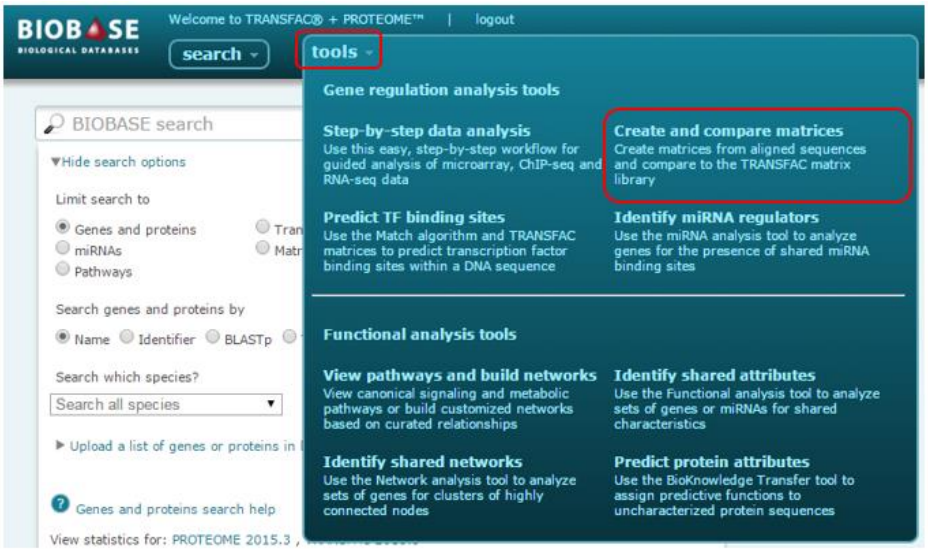


Finding novel motifs (Seeder)

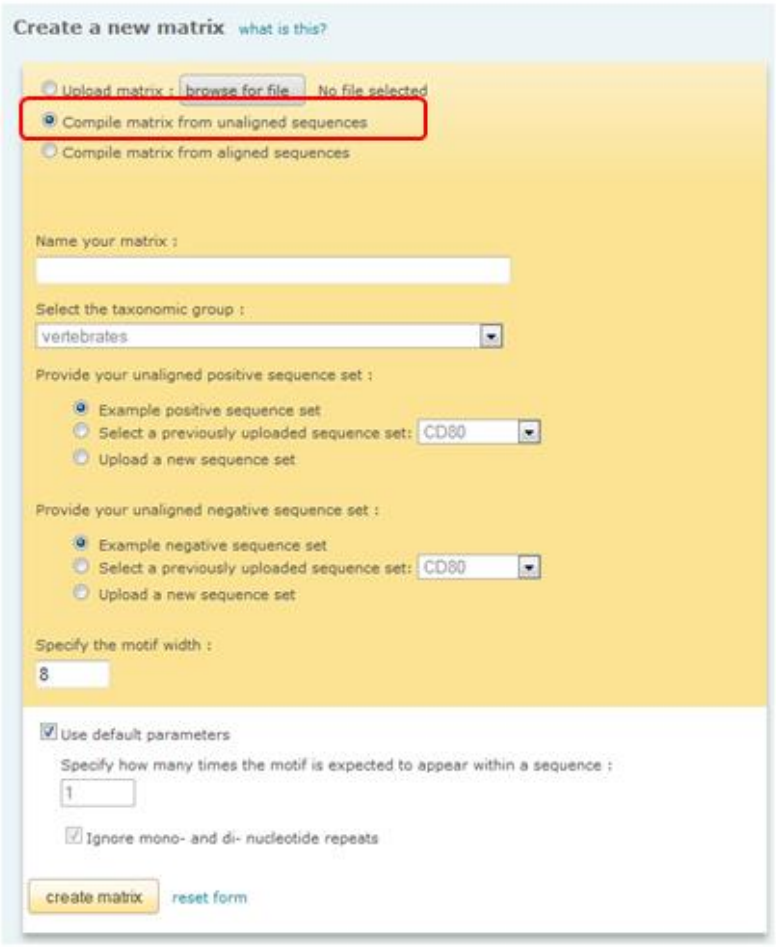
The DECOD algorithm in TRANSFAC replaces the Seeder algorithm ('**Analyze -> Binding sites search -> De-novo motifs (Seeder)**') used for finding novel motifs in a set of sequences in ExPlain:

For more information about this algorithm, please see the publication by Huggins et al. (2011) Bioinformatics 27:2361.

To identify a novel motif within a set of sequences, use the Create and compare matrices tools which is accessed from any page within TRANSFAC by clicking the tools menu followed by clicking the 'Create and compare matrices' link:


BIOBASE
BIOLOGICAL DATABASES
Welcome to TRANSFAC® + PROTEOME™ | logout
search
tools
BIOBASE search
Hide search options
Limit search to
Genes and proteins
miRNAs
Pathways
Search genes and proteins by
Name
Identifier
BLASTp
Search which species?
Search all species
Upload a list of genes or proteins in
Genes and proteins search help
View statistics for: PROTEOME 2015.3
Gene regulation analysis tools
Step-by-step data analysis
Use this easy, step-by-step workflow for guided analysis of microarray, ChIP-seq and RNA-seq data
Create and compare matrices
Create matrices from aligned sequences and compare to the TRANSFAC matrix library
Predict TF binding sites
Use the Match algorithm and TRANSFAC matrices to predict transcription factor binding sites within a DNA sequence
Identify miRNA regulators
Use the miRNA analysis tool to analyze genes for the presence of shared miRNA binding sites
Functional analysis tools
View pathways and build networks
View canonical signaling and metabolic pathways or build customized networks based on curated relationships
Identify shared attributes
Use the Functional analysis tool to analyze sets of genes or miRNAs for shared characteristics
Identify shared networks
Use the Network analysis tool to analyze sets of genes for clusters of highly connected nodes
Predict protein attributes
Use the BioKnowledge Transfer tool to assign predictive functions to uncharacterized protein sequences

Once the tool loads, select the 'Compile matrix from unaligned sequences' option and fill in the requested fields in the yellow section including the set of sequences to be searched for the novel motif, the set of background sequences to be used, and the expected motif width (a default setting of 8 is given):

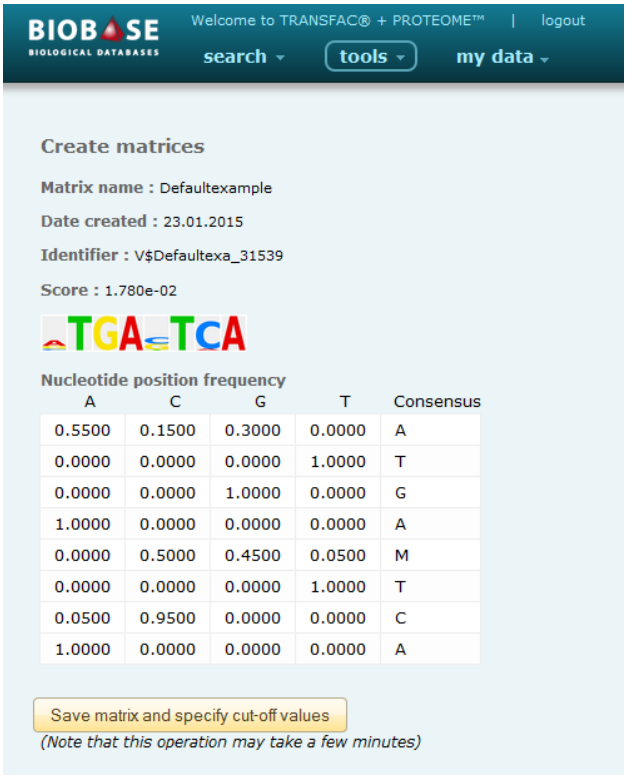

Create a new matrix what is this?
Upload matrix : browse for file No file selected
Compile matrix from unaligned sequences
Compile matrix from aligned sequences
Name your matrix :
Select the taxonomic group :
vertebrates
Provide your unaligned positive sequence set :
Example positive sequence set
Select a previously uploaded sequence set: CD80
Upload a new sequence set
Provide your unaligned negative sequence set :
Example negative sequence set
Select a previously uploaded sequence set: CD80
Upload a new sequence set
Specify the motif width :
8
Use default parameters
Specify how many times the motif is expected to appear within a sequence :
1
Ignore mono- and di- nucleotide repeats
create matrix
reset form

At this point you can start the analysis, keeping the remaining default parameters or by adapting them as desired in the white section:

1. Specify the number of times the motif is expected to appear within a sequence (default is set to 1)

2. Ignore mono- and di-nucleotide repeats (turned on by default)

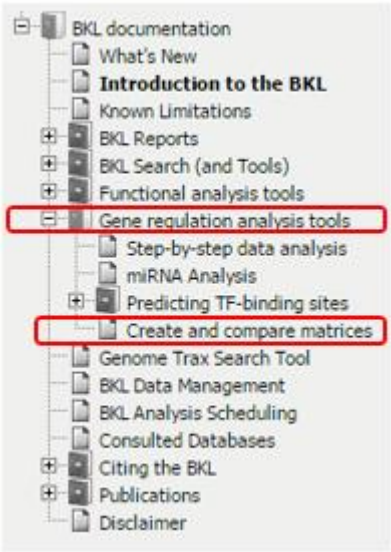
Once the analysis completes, a nucleotide position frequency table for the best scoring motif and consensus sequence logo preview will be created and you will be prompted to click the ‘save matrix and specify cut-off values’ button to finalize the upload:


BIOBASE
BIOLOGICAL DATABASES
Welcome to TRANSFAC® + PROTEOME™ | logout
search
tools
my data
Create matrices
Matrix name : Defaultexample
Date created : 23.01.2015
Identifier : V$Defaultexa_31539
Score : 1.780e-02
Nucleotide position frequency
A C G T Consensus
0.5500 0.1500 0.3000 0.0000 A
0.0000 0.0000 0.0000 1.0000 T
0.0000 0.0000 1.0000 0.0000 G
1.0000 0.0000 0.0000 0.0000 A
0.0000 0.5000 0.4500 0.0500 M
0.0000 0.0000 0.0000 1.0000 T
0.0500 0.9500 0.0000 0.0000 C
1.0000 0.0000 0.0000 0.0000 A
Save matrix and specify cut-off values
(Note that this operation may take a few minutes)

When the process is complete (this may take a few minutes due to the calculation of the cut-off values) you will be notified that the matrix has been saved to your library. It will now be listed in the 'Gene regulation analysis -> Data -> Matrices -> Uploaded matrices' folder of the 'my data' menu:

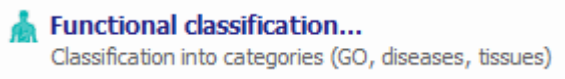
and will also be listed in the Profile creation tool (accessed from the Match tool page) which can be used to create a profile containing the matrix for use by Match.

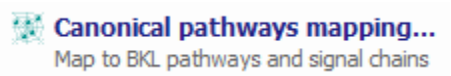
For a more detailed description of this tool, please see the 'Gene regulation analysis tools -> Create and compare matrices' page of the TRANSFAC user manual.



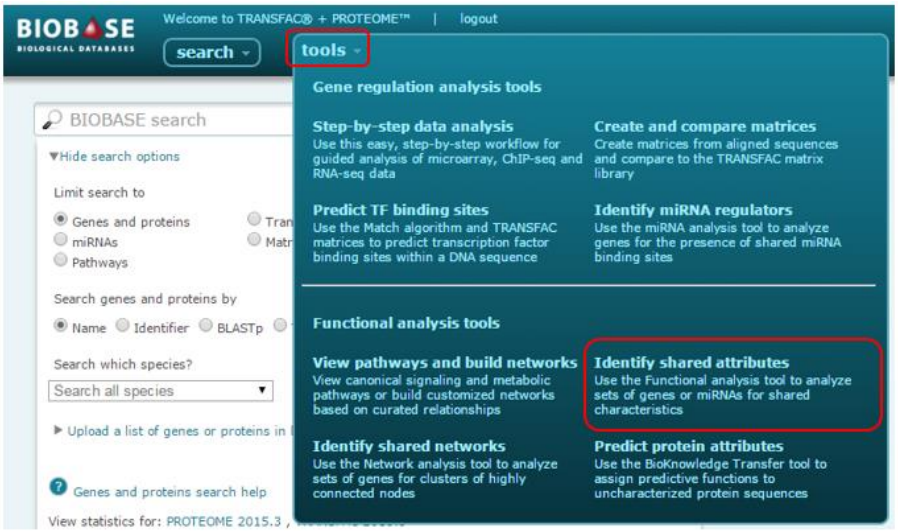
Functional analysis

The functional analysis tool in TRANSFAC replaces the '**Analyze -> Functional analysis -> Functional classification**' and '**Analyze -> Functional analysis -> Canonical pathways mapping**' functions of ExPlain:

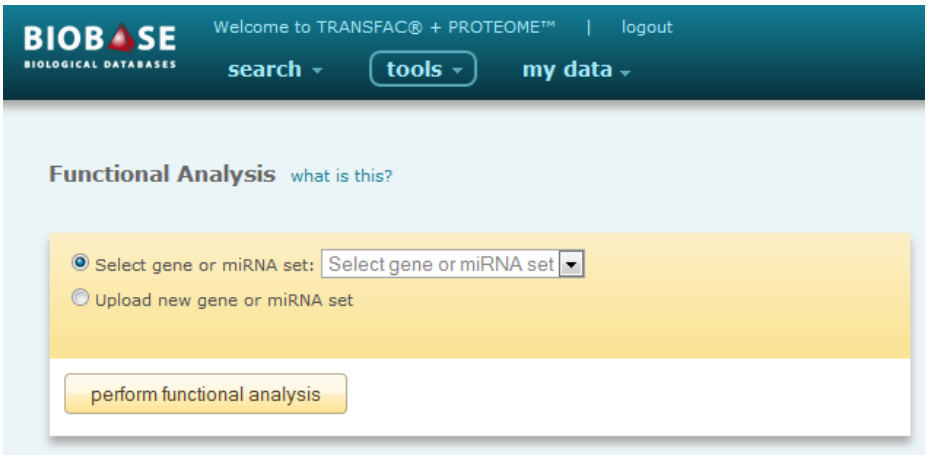




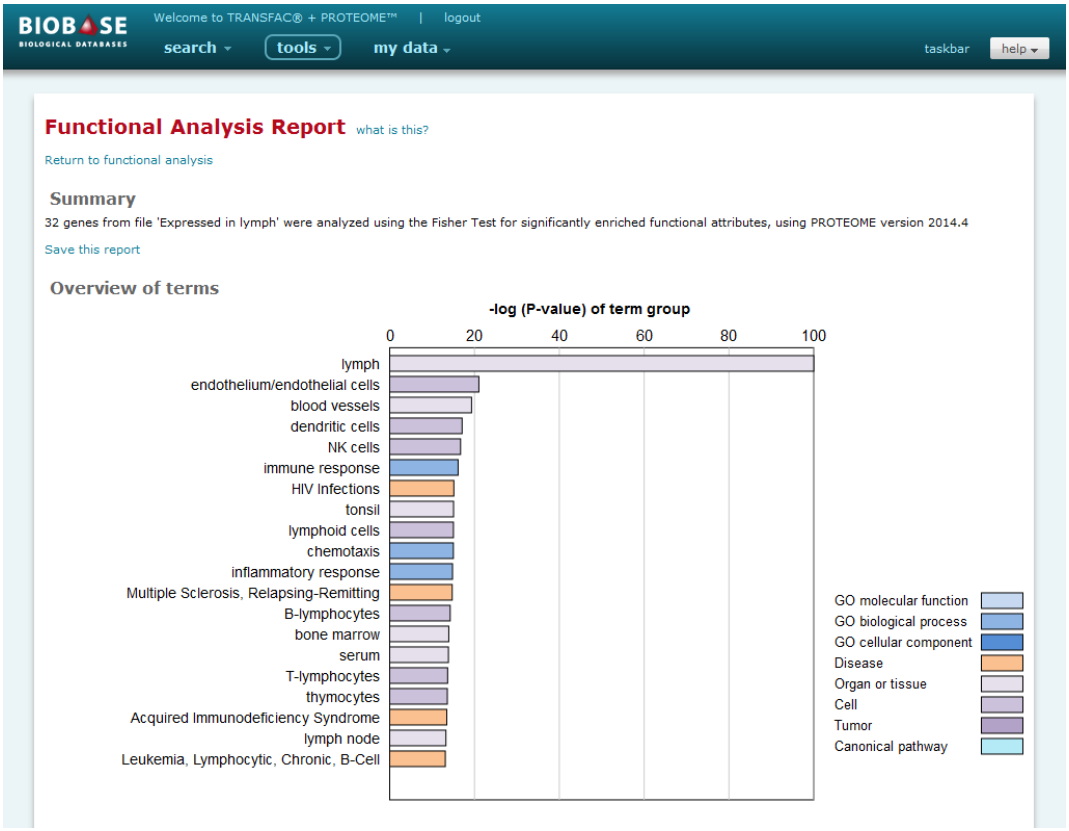
The tool analyzes sets of genes and miRNAs for the presence of statistically over-represented terms using a basic Fisher test analysis. To access the functional analysis tool, from any page within TRANSFAC, click the tools menu and then the 'Identify shared attributes' link:


BIOBASE
BIOLOGICAL DATABASES
Welcome to TRANSFAC® + PROTEOME™ | logout
search
tools
Gene regulation analysis tools
Step-by-step data analysis
Use this easy, step-by-step workflow for guided analysis of microarray, ChIP-seq and RNA-seq data
Create and compare matrices
Create matrices from aligned sequences and compare to the TRANSFAC matrix library
Predict TF binding sites
Use the Match algorithm and TRANSFAC matrices to predict transcription factor binding sites within a DNA sequence
Identify miRNA regulators
Use the miRNA analysis tool to analyze genes for the presence of shared miRNA binding sites
Functional analysis tools
View pathways and build networks
View canonical signaling and metabolic pathways or build customized networks based on curated relationships
Identify shared attributes
Use the Functional analysis tool to analyze sets of genes or miRNAs for shared characteristics
Identify shared networks
Use the Network analysis tool to analyze sets of genes for clusters of highly connected nodes
Predict protein attributes
Use the BioKnowledge Transfer tool to assign predictive functions to uncharacterized protein sequences
BIOBASE search
Hide search options
Limit search to
Genes and proteins
miRNAs
Pathways
Search genes and proteins by
Name
Identifier
BLASTp
Search which species?
Search all species
Upload a list of genes or proteins in
Genes and proteins search help
View statistics for: PROTEOME 2015.3

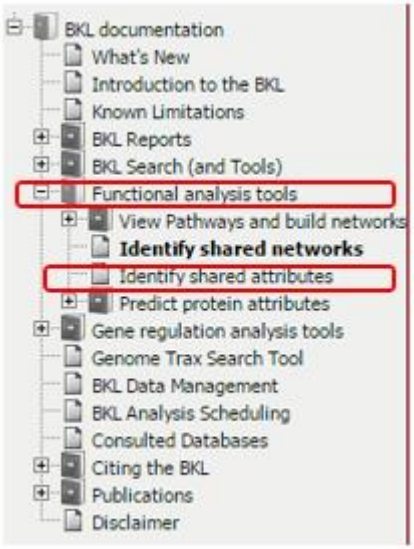
When the tool loads, you are given the option to analyze a previously uploaded list of genes or miRNAs or to upload a new list of genes or miRNAs. Despite the different entry point, the process for uploading a list of genes and miRNAs is the same as is described in [Loading gene and miRNA sets].



Once you have selected the data set to be analyzed, click the 'perform functional analysis' button. Your analysis will be forwarded to the taskbar. If you keep the taskbar window open until the analysis completes, the analysis results will automatically be loaded within the open window.

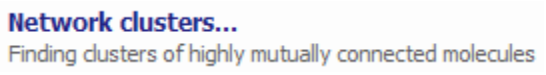


For a detailed description of the analysis report, please see the 'Functional analysis tools -> Identify shared attributes' page of the TRANSFAC user manual.

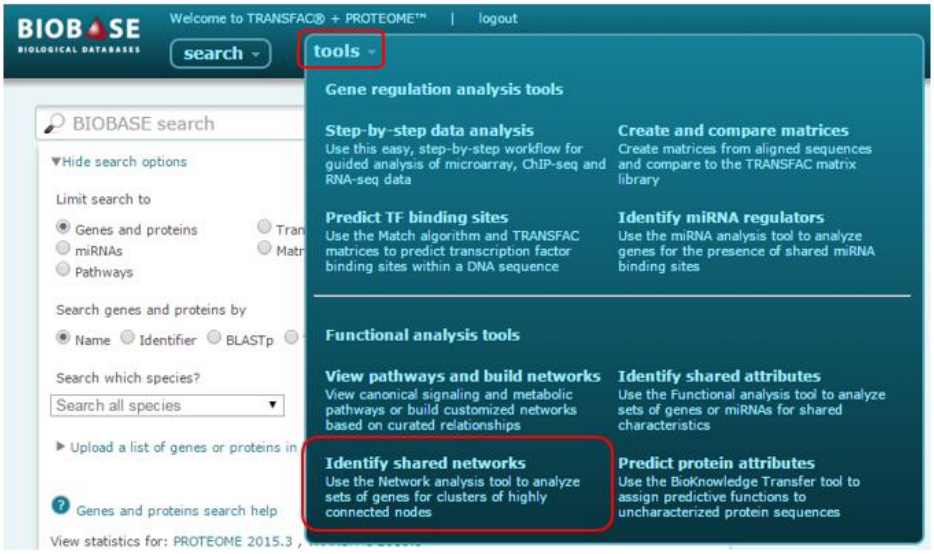


Network analysis

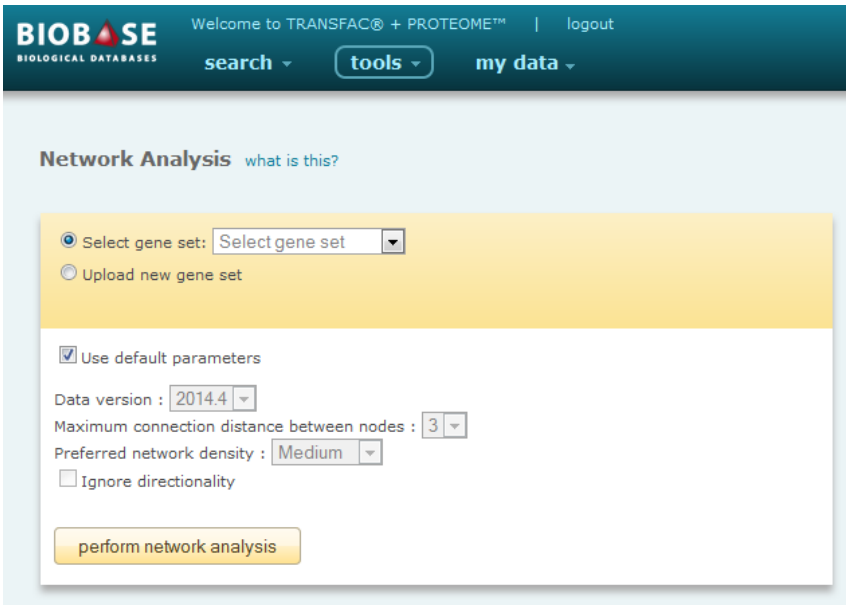
The network analysis tool in TRANSFAC replaces the '**Analyze -> Network analysis -> Network clusters**' tool of ExPlain:



The tool analyzes sets of genes for networks enriched with members of your gene set. To access the network analysis tool, from any page within TRANSFAC, click the tools menu and then the 'Identify shared networks' link:


BIOBASE
Welcome to TRANSFAC® + PROTEOME™ | logout
search
tools
Gene regulation analysis tools
Step-by-step data analysis
Use this easy, step-by-step workflow for guided analysis of microarray, ChIP-seq and RNA-seq data
Create and compare matrices
Create matrices from aligned sequences and compare to the TRANSFAC matrix library
Predict TF binding sites
Use the Match algorithm and TRANSFAC matrices to predict transcription factor binding sites within a DNA sequence
Identify miRNA regulators
Use the miRNA analysis tool to analyze genes for the presence of shared miRNA binding sites
Functional analysis tools
View pathways and build networks
View canonical signaling and metabolic pathways or build customized networks based on curated relationships
Identify shared attributes
Use the Functional analysis tool to analyze sets of genes or miRNAs for shared characteristics
Identify shared networks
Use the Network analysis tool to analyze sets of genes for clusters of highly connected nodes
Predict protein attributes
Use the BioKnowledge Transfer tool to assign predictive functions to uncharacterized protein sequences
BIOBASE search
Hide search options
Limit search to
Genes and proteins
miRNAs
Pathways
Search genes and proteins by
Name
Identifier
BLASTp
Search which species?
Search all species
Upload a list of genes or proteins in
Genes and proteins search help
View statistics for: PROTEOME 2015.3

When the tool loads, you are given the option to analyze a previously uploaded list of genes or to upload a new list of genes. Despite the different entry point, the process for uploading a list of genes is the same as is described in Loading gene and miRNA sets.



Once you have selected the data set to be analyzed, you can launch the analysis and it will be run with the default parameters which are set in the white section of the tool. If you wish to make any changes to the parameters, de-select the 'Use default parameters' and make the desired changes:

1. Data version

TRANSFAC supports the four most recent data versions. By default the current version is selected, but previous versions may be selected using the pull-down menu.

2. Maximum connection distance between nodes

This parameter specifies the maximum number of steps that may separate two nodes in the input list. By default the parameter is set to 3, but you may select a distance from 1 to 5. Specifying a smaller maximum connection distance will generally produce more, smaller networks while specifying a larger maximum connection distance will generally produce fewer, larger networks. In general, as you increase the maximum connection distance, smaller networks will become merged into larger networks.

3. Preferred network density

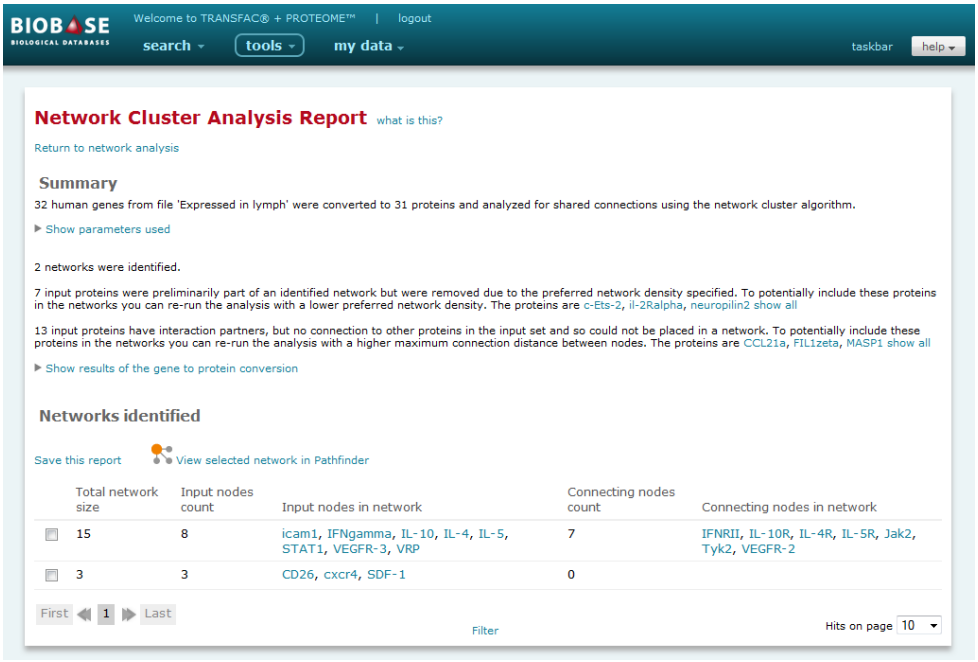
This parameter specifies the connectedness of a node to other nodes in the network. By default the parameter is set to Medium, but you may select densities of Very low, Low, Medium, High and Very

high. Specifying a lower preferred density favors the retention of nodes and will generally produce larger, more branched networks. Specifying a higher preferred density favors the removal of nodes whose connection to the network is more fragile and will generally produce smaller, more dense networks.

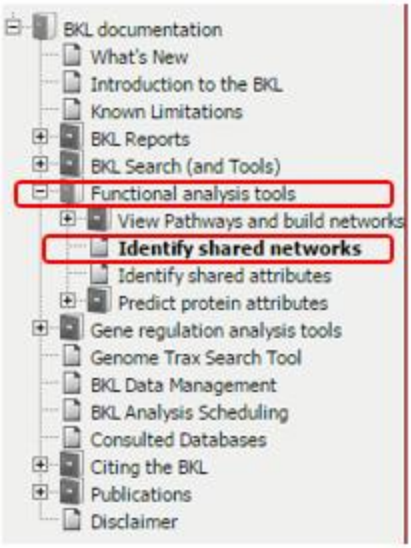
4. Ignore directionality

This parameter specifies the type of relationships that are considered when building the networks. By default, Ignore directionality is turned off and only those relationships which are unidirectional are considered. Examples of unidirectional relationships include a ligand activating its receptor, a kinase phosphorylating a target protein, etc. When Ignore directionality is turned off, the set of considered relationships is extended to also include bidirectional relationships. Examples of bidirectional relationships are protein-protein binding interactions which result in bidirectional complex formation. Ignoring directionality will generally produce larger networks and may merge smaller networks into larger networks.

When you are ready, launch the analysis by clicking the 'perform network analyis' button. Your analysis will be forwarded to the taskbar. If you keep the taskbar window open until the analysis completes, the analysis results will automatically be loaded within the open window.


BIOBASE
BIOLOGICAL DATABASES
Welcome to TRANSFAC® + PROTEOME™ | logout
search
tools
my data
taskbar
help
Network Cluster Analysis Report
what is this?
Return to network analysis
Summary
32 human genes from file 'Expressed in lymph' were converted to 31 proteins and analyzed for shared connections using the network cluster algorithm.
Show parameters used
2 networks were identified.
7 input proteins were preliminarily part of an identified network but were removed due to the preferred network density specified. To potentially include these proteins in the networks you can re-run the analysis with a lower preferred network density. The proteins are c-Ets-2, il-2Ralpha, neuropilin2 show all
13 input proteins have interaction partners, but no connection to other proteins in the input set and so could not be placed in a network. To potentially include these proteins in the networks you can re-run the analysis with a higher maximum connection distance between nodes. The proteins are CCL21a, FIL1zeta, MASP1 show all
Show results of the gene to protein conversion
Networks identified
Save this report
View selected network in Pathfinder
Total network size
Input nodes count
Input nodes in network
Connecting nodes count
Connecting nodes in network
15
8
icam1, IFNgamma, IL-10, IL-4, IL-5, STAT1, VEGFR-3, VRP
7
IFNRII, IL-10R, IL-4R, IL-5R, Jak2, Tyk2, VEGFR-2
3
3
CD26, cxcr4, SDF-1
0
First
1
Last
Filter
Hits on page 10

For a detailed description of the analysis report, please see the ‘Functional analysis tools -> Identify shared networks’ page of the TRANSFAC user manual.

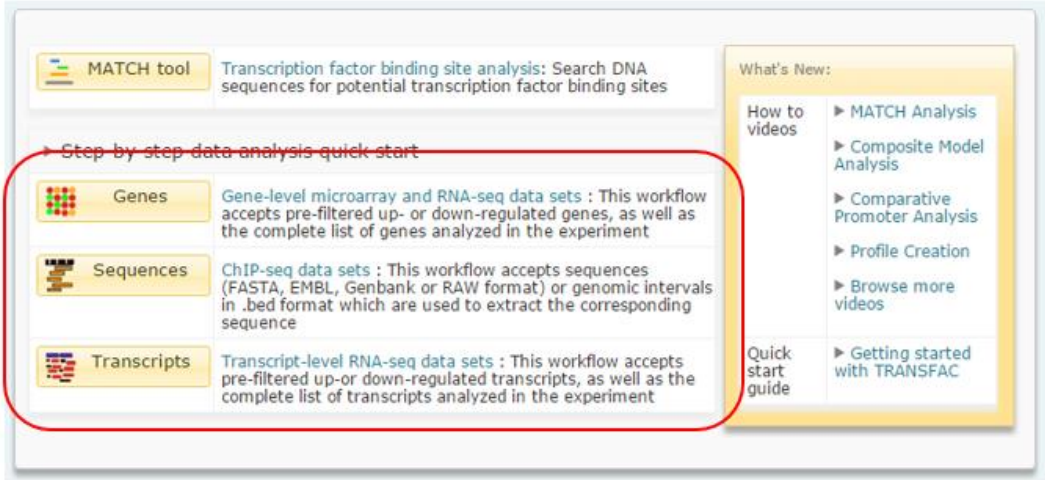


Step-by-step analysis (Wizard workflows)

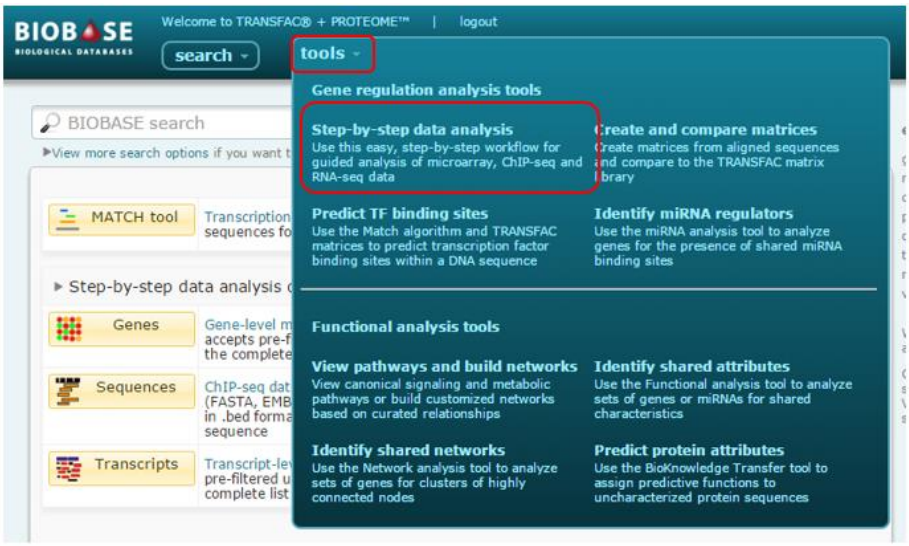
The step-by-step analysis tool in TRANSFAC replaces the ‘**Wizard mode**’ of ExPlain. Three workflows are supported:

1. Gene-level microarray and RNA-seq data sets
2. ChIP-seq data sets
3. Transcript-level RNA-seq data sets

Access the workflows through the quick start section of the Home page:


MATCH tool
Transcription factor binding site analysis: Search DNA sequences for potential transcription factor binding sites
Step-by-step data analysis quick start
Genes
Gene-level microarray and RNA-seq data sets : This workflow accepts pre-filtered up- or down-regulated genes, as well as the complete list of genes analyzed in the experiment
Sequences
ChIP-seq data sets : This workflow accepts sequences (FASTA, EMBL, Genbank or RAW format) or genomic intervals in .bed format which are used to extract the corresponding sequence
Transcripts
Transcript-level RNA-seq data sets : This workflow accepts pre-filtered up-or down-regulated transcripts, as well as the complete list of transcripts analyzed in the experiment
What's New:
How to videos
MATCH Analysis
Composite Model Analysis
Comparative Promoter Analysis
Profile Creation
Browse more videos
Quick start guide
Getting started with TRANSFAC

Or by clicking the tools menu and then the 'Step-by-step data analysis' link:


BIOBASE
Welcome to TRANSFAC® + PROTEOME™
logout
search
tools
Gene regulation analysis tools
Step-by-step data analysis
Use this easy, step-by-step workflow for guided analysis of microarray, ChIP-seq and RNA-seq data
Create and compare matrices
Create matrices from aligned sequences and compare to the TRANSFAC matrix library
Predict TF binding sites
Use the Match algorithm and TRANSFAC matrices to predict transcription factor binding sites within a DNA sequence
Identify miRNA regulators
Use the miRNA analysis tool to analyze genes for the presence of shared miRNA binding sites
Functional analysis tools
View pathways and build networks
View canonical signaling and metabolic pathways or build customized networks based on curated relationships
Identify shared attributes
Use the Functional analysis tool to analyze sets of genes or miRNAs for shared characteristics
Identify shared networks
Use the Network analysis tool to analyze sets of genes for clusters of highly connected nodes
Predict protein attributes
Use the BioKnowledge Transfer tool to assign predictive functions to uncharacterized protein sequences
BIOBASE search
MATCH tool
Genes
Sequences
Transcripts

For a detailed description of the workflow options, please see the 'Gene regulation analysis tools -> Step-by-step data analysis' page of the TRANSFAC user manual.

BKL documentation
What's New
Introduction to the BKL
Known Limitations
BKL Reports
BKL Search (and Tools)
Functional analysis tools
Gene regulation analysis tools
Step-by-step data analysis
miRNA Analysis
Predicting TF-binding sites
Create and compare matrices
Genome Trax Search Tool
BKL Data Management
BKL Analysis Scheduling
Consulted Databases
Citing the BKL
Publications
Disclaimer